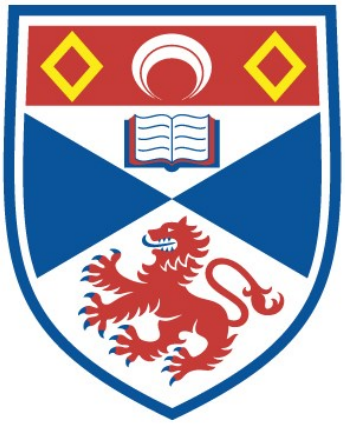
MUTAGENICITY OF 5-BROMOURACIL: QUANTUM CHEMICAL STUDY

Leo Holroyd

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2015

Mutagenicity of 5-bromouracil: Quantum chemical study

Leo Holroyd

University of St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD at the University of St Andrews

Date of Submission: 2 April 2015

Abstract

This thesis describes a computational investigation of the mutagenicity of 5-bromouracil (BrU).

In Chapter 1, three models of spontaneous and BrU-induced base mispairing (rare tautomer, wobble pair, and ion) are reviewed.

Chapter 2 presents the computational techniques used: electronic structure methods (Hartree–Fock-based and density functional theory) and molecular dynamics.

Chapter 3 presents optimisations of the keto and enol tautomers of BrU and uracil (U) in water clusters. The enol tautomer of BrU is found to be more stable than that of U.

Chapter 4 is a molecular dynamics study of the keto-enol tautomerism of BrU and U in a periodic water box. The pK_a of BrU at N3 is found to be lower than that of U.

Chapter 5 is a study of stacked base dimers containing BrU, U, or thymine (T) stacking with natural bases. Some structures were taken from the Protein Data Bank, while others were generated using an in-house methodology. BrU is found to stack more strongly than T *in vacuo*, but solvation and thermal effects nullify this difference.

Chapter 6 discusses the significance of the results in Chapters 3–5 in terms of BrU-induced mutagenesis.

Appendices A and B–D provide supplementary material to Chapters 2 and 5, respectively.

Appendix E is an investigation of the "base flipping" pathway of 2-aminopurine (2AP). Both 2AP/N and A/N dinucleosides (N = thymine or guanine) are found to adopt a wide range of energy-minimum conformations – not only stacked and "flipped", but also intermediate – and the stacked are not the most favourable by free energy.

Appendix F is a list of publications and papers in preparation. One publication concerns BrU stacking. The other is a conformational study of the dipeptide tyrosine-glycine: the theoretical results are shown to be consistent with experiment (R2PI spectra) if thermal effects are taken into account.

1. Candidate's declarations:

I, Leo Holroyd, hereby certify that this thesis, which is approximately 49,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in August 2011 and as a candidate for the degree of PhD in September 2012; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2015.

Date …… signature of candidate ……

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… signature of supervisor ……

3. Permission for publication: (*to be signed by both candidate and supervisor*)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY

a) Embargo on all of print copy for a period of two years on the following ground(s):
    - Publication would preclude future publication

Supporting statement for printed embargo request:

Some results in Chapter 5 and Appendix E form the basis of three papers currently in preparation to be submitted to peer-reviewed journals; we also intend to submit the results in Chapter 4 for publication in a peer-reviewed journal.

ELECTRONIC COPY

a) Embargo on all of electronic copy for a period of two years on the following ground(s):
    - Publication would preclude future publication

Supporting statement for electronic embargo request:

Some results in Chapter 5 and Appendix E form the basis of three papers currently in preparation to be submitted to peer-reviewed journals; we also intend to submit the results in Chapter 4 for publication in a peer-reviewed journal.

Date …… signature of candidate …… signature of supervisor ……

Please note initial embargos can be requested for a maximum of five years. An embargo on a thesis submitted to the Faculty of Science or Medicine is rarely granted for more than two years in the first instance, without good justification. The Library will not lift an embargo before confirming with the student and supervisor that they do not intend to request a continuation. In the absence of an agreed response from both student and supervisor, the

Head of School will be consulted. Please note that the total period of an embargo, including any continuation, is not expected to exceed ten years.
Where part of a thesis is to be embargoed, please specify the part and the reason.

Acknowledgements

This thesis would never have been possible without the ongoing personal and academic support of my supervisor, Dr. Tanja van Mourik, who has been unfailingly kind and patient throughout my studies. I have also depended heavily on the many other good people who have passed through Room 150 during my time here, and who have made it into a true community.

I have received particular technical and scientific assistance from Dr. Ragnar Björnsson, Prof. Michael Bühl, Mr. Luke Crawford, Dr. Herbert Früchtl, Prof. Marie-Pierre Gaigeot, Dr. Jan Götze, Prof. Anita Jones, Dr. James McDonagh, Dr. John Mitchell and Dr. Darren Smith. Furthermore, I would like to graciously thank everyone at the University of St. Andrews from whom I have benefitted through lectures, seminars, discussions, and trips to the pub.

Thanks must go to the Engineering and Physical Sciences Research Council (EPSRC) for providing financial support throughout my studies, and for allocating computer time *via* the National Service for Computational Chemistry Software (NSCCS). I am also deeply indebted to HPC-Europa 2 for funding my six-week placement at the University of Evry, France, with Prof. Marie-Pierre Gaigeot, and to everyone at Evry for their warm and generous welcome.

The calculations in this thesis were performed using three HPC clusters. Those in Chapters 3 and 5 were performed on the Wardlaw cluster at the University of St. Andrews, which is maintained heroically by Dr. Herbert Früchtl. Those in Chapter 4, and also some in Chapter 5, were performed on the Columbus and Slater machines of the NSCCS, and I would like to thank Drs. Ling Ge, Alexandra Simperler and Helen Tsui for their considerable assistance with them. While at Evry, I also performed calculations using the Jade supercomputer of CINES (Centre Informatique National de l'Enseignement Supérieur, France), but these were not finally presented in this thesis.

Table of contents

Abbreviations

2AP	2-Aminopurine
A	Adenine
a*n*Z (*n* = 2, 3, 4…)	aug-cc-pV*n*Z
BLYP	Becke-88 exchange, Lee–Yang–Parr correlation
B3LYP	Becke 3-parameter exchange, Lee–Yang–Parr correlation
BrU	5-Bromouracil
BSSE	Basis set superposition error
C	Cytosine *or* carbon
CBS	Complete basis set
CCSD(T)	Coupled cluster with singles, doubles and perturbative triples
CP	Counterpoise procedure
CPMD	Car–Parrinello molecular dynamics
-D (suffix)	Dispersion correction
df- (prefix)	Density fitting
DFT	Density functional theory
DNA	Deoxyribonucleic acid
d*X*MP (*X* = A, C, G, T, U or BrU)	Deoxy-*X*-osine/idine monophosphate
d*X*TP (*X* = A, C, G, T, U or BrU)	Deoxy-*X*-osine/idine triphosphate
E	Electronic potential energy
F	Helmholtz free energy
FTB	Face-to-back
FTF	Face-to-face
G	Guanine
G	Gibbs free energy
GGA	Generalised gradient approximation
H	Enthalpy
HF	Hartree–Fock
KS	Kohn–Sham
L (prefix)	Local orbital approximation
LIIC	Linear interpolation of internal coordinates
M05-2X	M05 exchange-correlation functional with double exact exchange
M06-2X	M06 exchange-correlation functional with double exact exchange
M06-L	Local M06 exchange-correlation functional
MD	Molecular dynamics
MP2	Møller–Plesset perturbation theory to 2nd order
mPW2PLYP	Modified Perdew–Wang exchange, Lee–Yang–Parr and MP2 correlation
NBO	Natural Bond Orbital
PBE0	Perdew–Burke–Ernzerhof exchange-correlation functional with exact exchange
PCM	Polarisable continuum model
PDB	Protein Data Bank
Pu	Purine
Py	Pyrimidine
QST*n* (*n* = 2 or 3)	Quasi-Newton synchronous transit-guided method

RI	Resolution of the identity
RNA	Ribonucleic acid
SCF	Self-consistent field
SCRF	Self-consistent reaction field
T	Thymine
TS	Transition state
U	Uracil
WC	Watson–Crick
ZPE	Zero-point energy

Chapter 1: Introduction

1.1. Structure of DNA

Deoxyribonucleic acid (DNA) is an essential component of all life on Earth. This compound, which is found in the nucleus of every cell, contains the genetic information necessary for growth and reproduction. Molecular biologists have established that the chemical and physical structure of the DNA molecule is the key to its functionality as an information carrier. Since the discovery of the celebrated double helix by Watson, Crick and others,[1-6] it has been understood that the storage and transmission of genetic information depends on the complementarity of two types of purine-pyrimidine base pair. This thesis therefore begins by introducing the key features of the DNA structure–function relationship. A discussion of mutations, especially in the context of base analogues, then follows.

The backbone of DNA consists of two helically interwoven strands of deoxyribose sugar rings, alternating with phosphate groups attached at the C'5 and C'3 positions. Each sugar is attached through its C'1 atom to one of four nitrogenous bases: the purines adenine (A) and guanine (G), which attach through their N9 atoms, and the pyrimidines thymine (T) and cytosine (C), which attach through N1. Each base is hydrogen-bonded ("base-paired") to a complementary base on the other strand: A pairs with T, and G pairs with C. The strands have directionality – one end of each strand, the 5' end, terminates with a phosphate, while the other, the 3' end, terminates with a hydroxyl. The strands are aligned in opposite directions, and their sequences (the DNA primary structure) are complementary, so that the canonical base pairs are always formed. The specific stereochemistry and hydrogen-bonding pattern of these pairs is known as Watson–Crick base pairing. A-T has two H-bonds, formed by N6H(A)•••O4(T) and N3H(T)•••N1(A), while G-C has three H-bonds, formed by N4H(C)•••O6(G), N1H(G)•••N3(C) and N2H(G)•••O2(C) (Figure 1.1).[7]

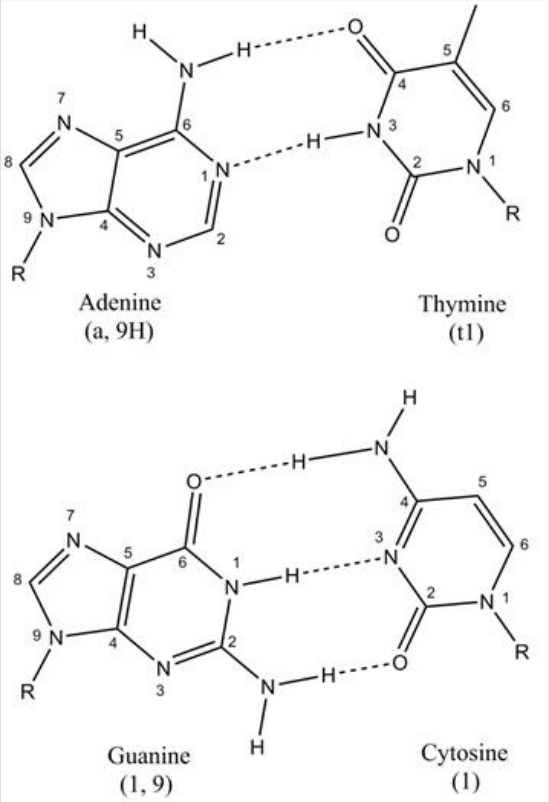


Figure 1.1: Watson–Crick A-T and G-C base pairs, featuring the canonical base tautomers (labelled in parentheses). "R" is the DNA sugar-phosphate backbone.

Note on terminology: correctly, *base* or *nucleobase* refers to a free molecule (e.g. molecular adenine); a base attached to a deoxyribose sugar is a *nucleoside* or *deoxyribonucleoside*, and is abbreviated with a small d (e.g. "dA" for deoxyadenosine); and a nucleoside bound to a phosphate is a *nucleotide* or *deoxyribonucleotide*, and is abbreviated with MP (e.g. "dAMP" for deoxyadenosine monophosphate).[8] In practice, because the backbone is chemically invariant for all nucleotides, it is common to refer to the units of DNA as "bases" (e.g. as adenine or A), even though each complete unit is in fact a deoxyribonucleotide monophosphate.

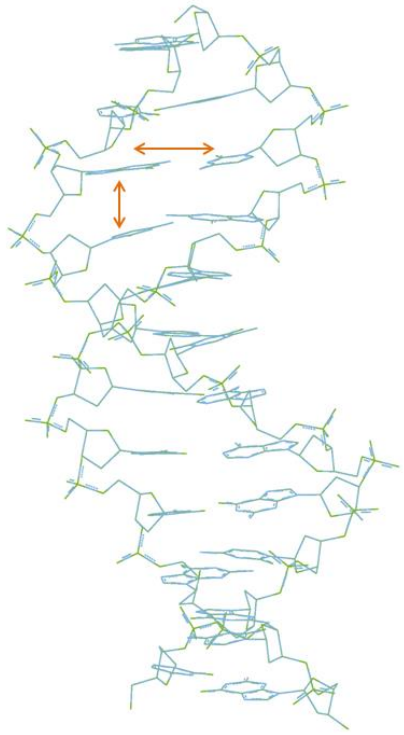


Figure 1.2: The classic double-helical shape of DNA, exemplified by the "Dickerson dodecamer" (PDB 1BNA).[6] The horizontal arrow indicates two paired bases; the vertical arrow indicates two stacked bases.

The four natural bases, or "letters" (A, C, G, and T), make up the genetic alphabet, in which each of the 21 protein-forming amino acids is coded for by a string of three letters, known as a codon. In this way, the DNA sequence of an organism encodes the amino acid sequences of its proteins. Since an alphabet of four letters allows 64 distinct three-letter sequences [$64 = 4^3$], this permits considerable degeneracy, with most amino acids represented by more than one codon. However, not all the DNA of an organism consists of codons – non-coding DNA accounts for a significant fraction of the genome, to an extent which varies between species.[9]

In the closely related molecule ribonucleic acid (RNA), which acts as an intermediary between DNA and proteins, the deoxyribose rings are replaced by ribose rings, and thymine is replaced by uracil (U), which has H rather than methyl at the 5-position. Hence the genetic alphabet of RNA is A, C, G and U. The global structure of RNA is complicated, and quite different from that of DNA,[10] and it will not be covered here.

The specificity of each natural base to pair exclusively with its correct partner, giving A-T and G-C as the only "correct" combinations, is the essence of the information content of DNA, and also the reason for the stability of that information against decay. When DNA is replicated during cell division, the two strands are unwound and each acts as a template for the synthesis of a new strand. Each template base is matched with the complementary incoming base (known as the substrate), so that, for example, the presence of template A always results in substrate T. Thus the sequence is maintained through each round of replication.

While the two strands are bound to each other by planar, hydrogen-bonded base pairing, another interaction, known as base stacking, operates vertically within each strand. The helical secondary structure of DNA permits considerable overlap and close contact between the planar aromatic rings of sequential bases, and this is energetically favourable overall. The subtle balance of attractive and repulsive forces that contribute to stacking (see Chapter 5) includes dispersive, electrostatic, charge-transfer, exchange, polarisation and entropic effects; it is also heavily influenced by the presence of water, resulting in both hydrophilic and hydrophobic interactions.[11-17] While base *pairing* embodies the information content of DNA, and provides its self-replicative ability, base *stacking* is believed to be the dominant source of the thermodynamic stability of DNA; in fact, base pairing may be on balance energetically destabilising.[18]

A sequence of two consecutive base pairs is known as a *base step*. Base steps contain two planar base pairing interactions, and two vertical base stacking interactions. For example, the sequence 5'-TC-3' on the left strand will be matched by the sequence 5'-GA-3' on the right strand, and the corresponding step is denoted TC:GA. The two hydrogen bonding interactions in this step are T-A and C-G, and the two stacking interactions are T/C and G/A. We typographically distinguish between hydrogen bonding and stacking by denoting planar base pairs with a hyphen, e.g. T-A, and vertical stacked dimers with a forward slash, e.g. T/C.

1.2. Errors and mutations

The fidelity of correct base pairing is extremely high, but not perfect: errors occur when the wrong incoming base is incorporated into the template strand, leading to a pair other than A-T or G-C. These mispairs occur spontaneously at an estimated rate of around 10^{-3} to 10^{-5} per base pair during replication,[19] and are a major source of substitution mutations – permanent, heritable changes in DNA sequence, involving the replacement of one letter with another. (Other kinds of mutation exist but will not be discussed here.) Substitutions involving the replacement of a pyrimidine with another pyrimidine, or a purine with another purine, lead to A-T→G-C or G-C→A-T pathways, and are called *transitions*. Substitutions involving the replacement of a pyrimidine with a purine, or *vice versa*, lead to A-T→T-A or G-C→C-G pathways, and are called *transversions*. Both are generally deleterious when they occur in coding sections of the genome, so cellular machinery exists to repair mismatches before they are propagated as mutations. Due to the action of repair enzymes, the spontaneous mutation rate in living organisms is considerably lower than the replicational error rate.[20] Spontaneous mutation rates vary widely between species, and also between different regions of the genome within a single species, being especially high at certain "hot spots"; this being taken into account, they are generally within the range of 10^{-7} to 10^{-11} per base pair.[19-24] Mutations are the raw material of evolutionary change.

1.3. Structure of mispairs

The cause of replicational errors has been much debated ever since the discovery of complementary base pairing. In the paper which introduced the double helix, Watson and Crick suggested the involvement of rare tautomers of the natural bases.[5] The hydrogen atoms on each base can in principle adopt several different locations, giving rise to a variety of possible tautomers. In practice, a single tautomer of each base – the one required for canonical base pairing – heavily predominates under physiological conditions. However, Watson and Crick proposed that rare tautomers may occasionally arise, which mimic the hydrogen bonding profile of a different base. In their example, A may adopt a rare imino form (systematically labelled "i1, 1H,9H"[25] but denoted A* herein).[i] The H-bonding profile of this tautomer is equivalent to G, so it can base-pair in normal ("pseudo-Watson–Crick") geometry with C but not with T. If this occurs during replication, one of the two progeny strands will feature an A*-C base pair. When *this* progeny strand replicates, C will pair in normal fashion with G, creating a G-C base pair in one of its own two progeny strands. Thus the A*-C mispair behaves as the intermediate in an AT→GC transition. (The other progeny strand formed from A*-C will contain the correct A-T pair.)

[i] An asterisk (*) herein indicates any rare base tautomer that is potentially involved in mispairing.

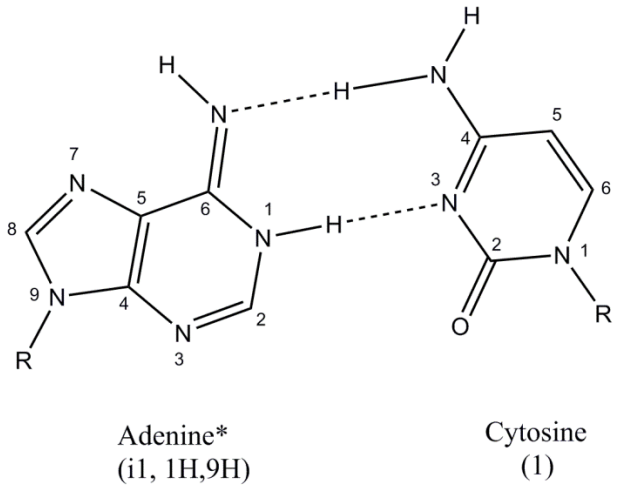


Figure 1.3: Mutagenic base pair between the rare imino tautomer of adenine (A*) and the canonical tautomer of cytosine (C), in Watson–Crick geometry.

In an elaboration of this theory, Löwdin proposed that mutagenic rare tautomers arose through *double quantum tunnelling* of the protons involved in hydrogen bonding.[26] Here, the wavepackets of one H-bonded proton on each base simultaneously tunnel into the shallower potential wells associated with covalently bonding to the other base. This intermolecular proton exchange results in rare tautomers of both bases, giving either G*-C* or A*-T* base pairs. Both bases are then expected to mispair upon replication. Löwdin's elegant model continues to raise considerable theoretical and practical interest as a case study in the emerging interdisciplinary field of "quantum biology".[27] However, most attention has focused on the role of water in intramolecular proton transfer, causing tautomerisation of a single base, as per Watson and Crick.[28]

Although attractive and plausible, the "rare tautomer hypothesis" has eluded direct experimental verification. For practical reasons, it is difficult to obtain nucleic acid structures during the replication process with sufficient atomic resolution to unambiguously identify a particular tautomer. Most of the support for this model since its original proposal has instead been circumstantial. Topal and Fresco noted an approximate agreement between the observed mutation frequency in a particular gene of *E. coli* and the *square* of the estimated rare tautomer concentration of the natural bases; they reasoned that the interval during the two-step replication process (synthesis followed by checking) allowed rare tautomers a chance to return to the major tautomer before the process was complete, thus reducing the final error rate by a factor of itself.[21] Sinha and Haimes noted that forced transition mutations during DNA synthesis proceeded almost entirely through G-T rather than A-C intermediates, and proposed that this was because – within the rare tautomer model – G-T mispairs are more stable, since they contain three hydrogen bonds compared to only two in A-C.[29]

However, Morgan published a sceptical review of the rare tautomer hypothesis, challenging the general acceptance in the secondary literature of tautomerism as a source of mutations.[30] The author noted that the tautomeric ratios of nucleobases are difficult to measure with quantitative accuracy, and cannot be directly extrapolated to nucleotides in DNA. As a rival to this model, some groups have focused on mispairs that involve both bases in their major tautomeric forms. The major tautomers of G and T can form an incorrect base pair with two hydrogen bonds, but the bases are shifted out of normal Watson–Crick geometry due to the requirement of using different donor and acceptor sites from the G-C pair. The resulting base pair is equivalent to the "wobble pairs" which Crick realised were crucial in RNA as the source of the degeneracy of the genetic code,[31] but which are not usually

found in DNA. G-T wobble pairs certainly exist, and have been shown to be stable in synthetic crystalline DNA.[32-34] A-C can also exist as a wobble pair, although one base must be protonated (positively charged) to allow two hydrogen bonds to form.[35, 36] Stable mispairs can thus be incorporated into DNA without the formation of rare tautomers.

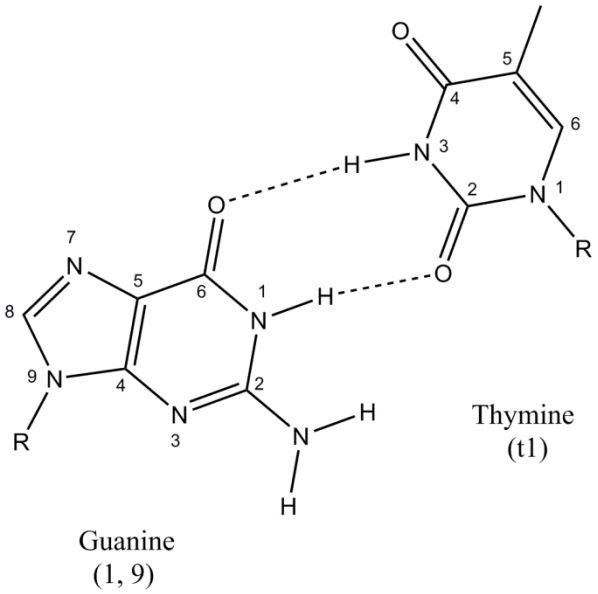


Figure 1.4: Mutagenic base pair between the canonical tautomers of guanine and thymine, in wobble geometry.

However, the role of wobble pairs in mutation has been strongly questioned. Strazewski and Tamm, while conceding that mutation rates rarely correspond to measured tautomeric equilibria – and that *average* mutation rates are in any case misleading, since mutation is strongly dependent on nucleotide pool concentration, sequence, and polymerase – nevertheless asserted that the correct stereochemistry, which wobble pairs do not possess, is crucial in passing undisturbed through the replication process.[37] If this is so, wobble pairs should be much more easily recognised and removed by the cellular repair machinery than rare tautomeric mispairs in Watson–Crick geometry. Casting doubt on the notion that crystallographic observation of wobble pairs implies that they are the main cause of mutation, they concluded that the formation of such pairs in cellular DNA should be rare, and could only account for a low, non-site-specific "background" rate of mutation that does not explain the existence of hot spots.

Developing this idea, Suen *et al.* argued that, since replication does not take place under conditions of thermodynamic equilibrium, the repair apparatus cannot be expected to discriminate between correct and incorrect base pairs on thermodynamic grounds alone, notwithstanding the higher energy of disfavoured tautomers.[38] Instead, they advocated a role for stereochemical selection. Here they cited Moran *et al*'s finding that difluorotoluene, an isosteric analogue of T which is unable to hydrogen-bond with A, nevertheless retains a high specificity of incorporation opposite A in Watson–Crick geometry, solely by virtue of its steric resemblance to T.[39] Defending the role of rare tautomers in mutagenesis, Suen *et al.* provided spectroscopic evidence that the mutagenic cytosine analogue 5-hydroxy-2'-deoxycytidine has a higher abundance of the rare imino tautomer – which mimics T and can therefore mispair with A – than C itself does.[38] They attributed the increased stability of the imino tautomer to electron donation from the ionised 5-hydroxyl group, which is not present in cytosine. Similarly, while studying the mutagen 6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-6H,8H-pyrimido-[4,5-c][1,2]oxazin-2-one, Harris *et al.* reported an extremely close agreement between the ratio of imino and amino tautomers of this compound and the ratio of its incorporation opposite A and

G, respectively, during DNA replication.[40, 41] Noting that the imino-A and amino-G base pairs both have Watson–Crick geometry, they argued for selection on stereochemical rather than thermodynamic grounds. Their experimental tautomeric ratio, measured by spectrophotometry, was subsequently endorsed by the theoretical calculations of Podolyan *et al*.[42]

The rare tautomer hypothesis has also attracted recent support from X-ray diffraction studies. Khuu and Ho observed a highly wobbled A-T base pair in a four-stranded DNA complex known as a Holliday junction, and attributed the distortion of Watson–Crick geometry to an enol tautomer of T stabilised by four contiguous G-C base pairs.[43] They suggested that such stabilisation of rare tautomers by G-C contributes to the relatively high rate of transition mutations in G-C rich sequences of DNA.

In 2011, direct evidence of rare tautomer involvement in mispairing was provided for the first time by two studies. Bebenek *et al*. performed a crystallographic study of G-T mismatches, containing G as the incoming base paired with template T.[44] The misinsertions were catalysed by a derivative of human DNA polymerase λ, differing from the wild type enzyme by the deletion of five residues, an alteration which was found to increase the rate of mutagenic mispairing. When located at the polymerase active site – i.e. immediately after misinsertion – the mispair had Watson–Crick geometry. However, when located at the 3' terminus of the DNA strand, the mispair had wobble geometry. The authors suggested that initial misinsertion events occur in Watson–Crick geometry, but are subsequently resolved into wobble pairs before being propagated as mutations. They also proposed that there is a common catalytic mechanism for the insertion of correct and incorrect nucleotides, given their structural similarity. However, the precise nature of their Watson–Crick G-T mispair was unclear – the O6(G)-O4(T) distance was consistent with an H-bond, implying tautomerisation of one base, but the rate of misinsertion was found to increase with pH, implying an ionised mispair with only two H-bonds (see also below, Section 1.4.).

In the same year, Wang *et al*. studied A-C mispairs in the insertion site of a DNA polymerase that catalyses replication in crystals.[36] In the presence of Mg^{2+}, A-C formed a wobble pair, but in the presence of Mn^{2+} (a known mutagen), A-C adopted a pseudo-Watson–Crick geometry that is only possible following tautomerisation (not ionisation) of either A or C. The rare-tautomeric A-C pair was found to be stabilised by a hydrogen bond with water that has an equivalent in all cognate base pairs (water being naturally present at the surface of cellular DNA[45]), but which is geometrically hindered in the wobble pair. They also proposed that the polymerase may alter the tautomeric constant in favour of the rare tautomeric pair, and that Mn^{2+} binds to cognate base pair shapes more strongly than Mg^{2+}, and therefore promotes the tautomeric variant of A-C. The role of coordinating metal ions in stabilising rare base tautomers may in fact be extensive,[46] although it will not be further examined here.

It is important to note that the equilibrium population of the "mutagenic" rare tautomer is different for each natural base. In a comprehensive set of studies, Hanus *et al*. determined the relative energies of all the possible tautomers of A, C, G, T and U by *ab initio* methods: in the gas phase, with monohydration, and with continuum solvent.[25, 47-49] The portion of those results relevant to base mispairing is summarised in Table 1.1. It can be seen that, in the gas phase, the mutagenic tautomers of C and G are considerably more stable than those of A, T and U, and that although hydration does reduce this difference, it still appears to be energetically significant in water. The calculated aqueous value of ΔG is 5.2 kcal/mol for cytosine and 5.7 kcal/mol for guanine. These values match the

experimental estimates of the population ratios, K_t, of the canonical tautomers to the second most abundant tautomers, which range from 10^4 to 10^5.[46] ΔG and K_t are related by

$$\Delta G = -RT \ln K_t \quad (1.1)$$

Setting T = 298 K, the lower estimate of K_t (10^4) gives ΔG = 5.5 kcal/mol, while the upper estimate of K_t (10^5) gives ΔG = 6.8 kcal/mol.

Base	$\Delta E_{tautomer}$ (gas phase)	$\Delta E_{tautomer}$ (monohydrated)	$\Delta G_{tautomer}$ (aqueous)
Adenine	12.07	10.97	8.01
Cytosine [a]	1.31	2.57	5.2
Guanine	0.09	1.44	5.69
Thymine	13.35	10.02	9.78
Uracil	12.54	9.32	8.81

Table 1.1. Energies in kcal/mol of the mutagenic tautomers of the natural bases, relative to the canonical tautomers, calculated by Hanus *et al.*[25, 47-49] Potential energies (ΔE) were calculated with RI-MP2/TZVPP. Free energies in aqueous solution (ΔG) were evaluated at 298 K by the molecular dynamics–thermodynamic integration method.

[a] The canonical (amino-oxo) tautomer of cytosine (labelled "1" in Figure 1.1) is *not* the most stable in the gas phase, or with monohydration. The most stable under these conditions is the right-handed amino-enol tautomer (labelled "2a" by Hanus *et al.*;[47] not shown here). However, the gas-phase and monohydrated energies of the mutagenic (left-handed imino-oxo) tautomer of cytosine (labelled "3b" in Figure 1.5) are nonetheless given here relative to the canonical tautomer. (In aqueous solution, the canonical tautomer is the most stable.)

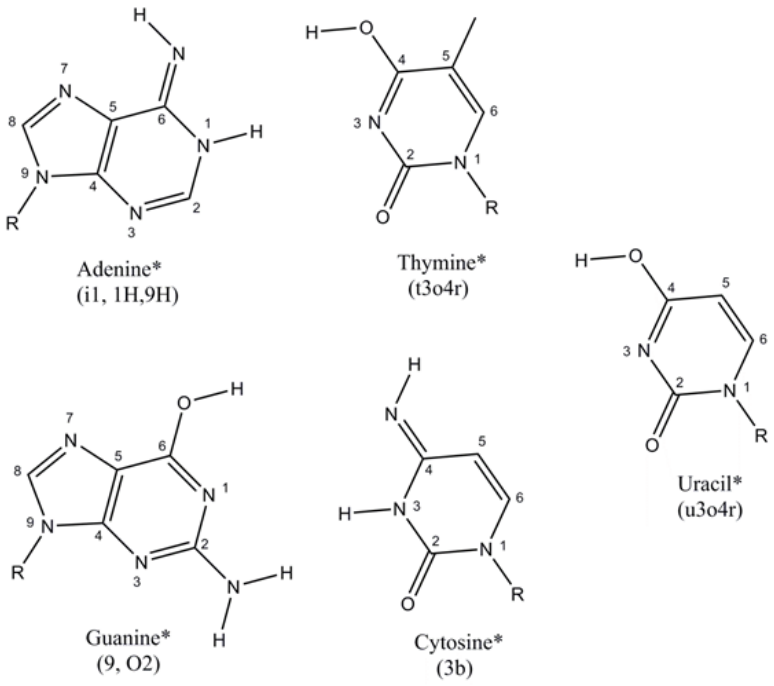


Figure 1.5: The mutagenic tautomers of the natural bases (including uracil) considered in Table 1.1.

The group of Leszczynski has also been highly active in determining, by theoretical methods, the equilibrium populations of rare base tautomers, as well as the stability of rare tautomeric mispairs, the effects of hydration and the DNA backbone, and potential mechanisms of single and double proton

transfer in nucleobases. In surveying the literature, they summarised that the rare tautomeric populations of G and C are considerably higher than those of A and T (in agreement with Hanus *et al.*, cf. Table 1.1), and actually exceed the observable frequency of spontaneous mutations. They therefore support the rare tautomer hypothesis.[50] Another recent review of the hypothesis, focusing on theoretical studies, was provided by Brovarets' *et al.*[28] From their own results, the authors determined that the A-C* and G*-T mispairs have lifetimes, dynamical stabilities, dissociation energies and equilibrium populations that are consistent with the rare tautomer model of spontaneous mutagenesis. However, they ruled out the Löwdin mechanism of intermolecular double proton transfer, on the basis that the A*-T* mispair is not a minimum on the Gibbs free energy surface with MP2/6-311++G(2df,pd)//B3LYP/6-311++G(d,p) (repeating an earlier finding by Florián and Leszczynski),[51] while the lifetime of the G*-C* mispair is too short to allow its involvement in replication.

Both groups therefore concluded that single base substitutions ("point mutations") arising from the formation of rare tautomers should primarily involve the tautomerisation of C and G, rather than A or T. Assuming the rare tautomers are located in the template strand, this will result in GC→AT transitions. Florián and Leszczynski argued that this mechanism implies that the GC content of DNA should decrease over evolutionary time, and noted in support of it that the GC content of higher organisms is indeed markedly lower than the AT content.[51]

1.4. Mutagenicity of 5-bromouracil: Experimental knowledge

Besides the four natural bases, DNA can also incorporate other molecules in their place. These *base analogues* can have important effects both on the physical properties of DNA and on its information content (i.e. genetic effects).

5-Bromouracil (BrU) is an analogue of thymine, with the 5-methyl group replaced by bromine. For decades it has been known to be highly mutagenic, i.e. it increases the mutation rate of DNA well above the spontaneous rate.[52-57] The induced mutations are exclusively of the transition kind. BrU is widely used in the laboratory as an experimental mutagen, but it can also be created physiologically (from uracil) by eosinophil peroxidase, a specialised cytotoxic protein used by anti-parasitic white blood cells.[58]

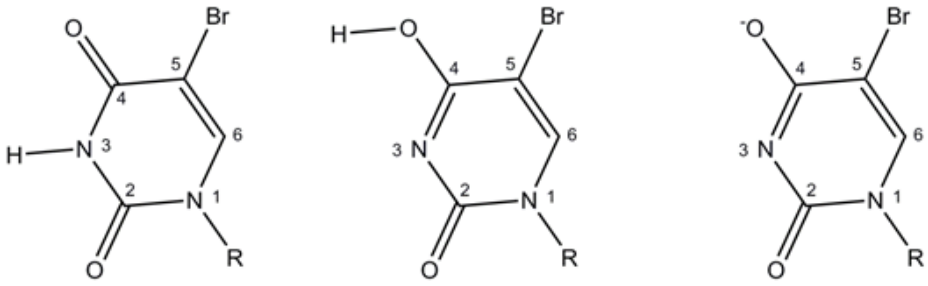


Figure 1.6: Three tautomers of 5-bromouracil.
Left: canonical (diketo). Centre: "mutagenic" (O4-enol). Right: anionic.

Since the 1950s, the mutagenicity of BrU has generally been explained in terms of mispairing and misincorporation.[54-56] Upon exposure to cellular DNA, BrU is incorporated in place of its parent

compound T – which it resembles sterically – at a replacement rate that can be varied widely, up to nearly 100%.[59] Upon replication, however, the incorporated BrU is more susceptible than T to mispairing with G, causing AT→GC transitions just as the T-G mispair does. Alternatively, BrU may be present in the nucleotide pool used for growing the new strand, and an incoming BrU could misincorporate opposite G in the template, leading eventually to a GC→AT transition.

Hence BrU is mutagenic because, although it can be stably incorporated in place of thymine, it has a lower specificity for pairing with adenine. In the 1970s and '80s this explanation was for a time challenged by the observation that the rate of BrU-induced mutagenesis could be controlled without varying the rate of BrU incorporation into DNA. Kaufman, Davidson and co-workers proposed that BrU, when exposed to the cell, caused mutations by unbalancing the pool of nucleotides available for the growing strand: specifically, decreasing the ratio of C to the combined amount of BrU and T below some mutagenic threshold.[60, 61] This "deoxycytidine starvation" apparently induced mutations by some unclear mechanism that depended on the level of exposure to BrU, but did not require direct incorporation of BrU into DNA. Later results from the same group, however – which did establish a correlation between mutagenesis and BrU incorporation – led them to conclude that mispairing, stimulated by nucleotide pool imbalance, *was* the mechanism of BrU-induced mutagenesis.[62, 63] Mispairing of template BrU with incoming G, stimulated by a high ratio of G to A in the pool, caused AT→GC transitions ("DNA-dependent mutation"), while the earlier case involving deoxycytidine starvation was attributed to mispairing of template G with incoming BrU, leading to GC→AT transitions ("pool-dependent mutation"). Nucleotide pool imbalance was determined to be a necessary but not sufficient condition for inducing mutations.[64]

The role of BrU mispairing is now generally accepted. However, it remains unclear *why* BrU is more susceptible to mispairing than T is. Like the T-G mispair, the structure of the BrU-G mispair has not been definitively established. The oldest and most widely accepted explanation – based on Watson and Crick's rare tautomer hypothesis, and suggested by Crick, among others, for this particular mutagen – is that BrU has a greater abundance of the O4-enol tautomer than T does.[21, 56, 65] This tautomer (systematically labelled "bru3o4r"[66] but denoted BrU* herein), which can be formed by the movement of a proton from N3 to O4, mimics the hydrogen bonding profile of C, and can therefore mispair with G in pseudo-Watson–Crick geometry (Figure 1.7), just like the equivalent tautomer of T can.

Early experimental support for this hypothesis was provided by Katritzky and Waring, who measured a greater tautomeric constant for the enol tautomer of 1-methyl-5-BrU than for 1-methyluracil in aqueous solution, using the basicity method.[65] However, Morgan was sceptical of the relevance of this experiment to mispairing *in vivo*, noting that the tautomeric constants may have been strongly affected by both methylation and the highly acidic conditions necessary for the experiment.[30] Thymine and uracil are predicted to have very low populations of their "mutagenic" tautomers compared to cytosine and guanine,[49] so the inclusion of a bromine atom would need to have a large effect on the tautomeric ratio to be experimentally measurable. The rare tautomer hypothesis for BrU has become standard in the teaching literature,[20, 23, 67-69] but no direct evidence of an enolised BrU*-G mispair in a DNA helix has yet been published.

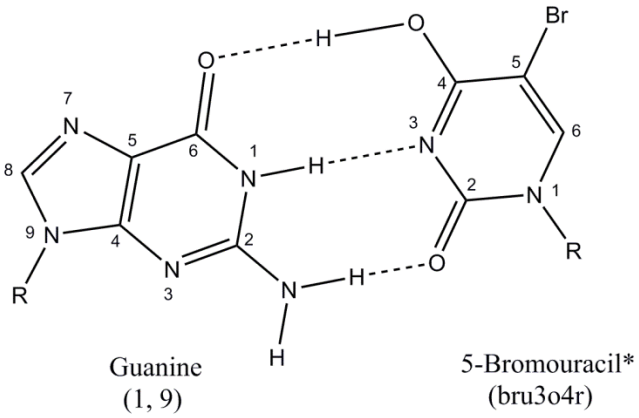


Figure 1.7: Mutagenic base pair between the canonical tautomer of guanine and the O4-enol tautomer of 5-bromouracil, in Watson–Crick geometry.

A rival explanation is the wobble pair model. The canonical keto tautomer of BrU can base-pair with G *via* two hydrogen bonds, just as T can (Figure 1.8). BrU-G wobble pairs have been observed by X-ray diffraction in short synthetic DNA polymers.[34, 70, 71] A wobble pair has also been observed between guanine and the chemotherapeutic base analogue 5-fluorouracil (FU).[72] Within the wobble pair model, the greater abundance of BrU-G mispairs, compared to T-G, is usually explained in terms of stronger stacking interactions – if BrU stacks more strongly than T, its mispairs are less likely to be repaired during replication, whatever stereochemistry they have.[70, 73]

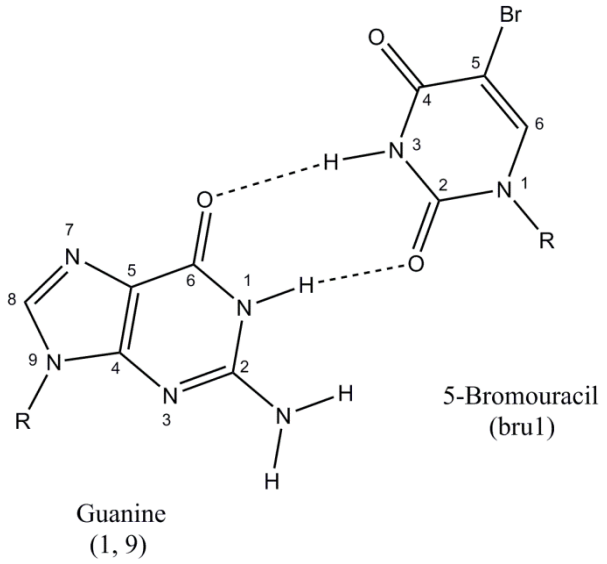


Figure 1.8: Mutagenic base pair between the canonical tautomers of guanine and 5-bromouracil, in wobble geometry.

This model has the same strengths and weaknesses as that of wobble pairs involving the natural bases – it is consistent with *in vitro* crystallographic results, and obviates the need for a rare tautomerisation event, but has been criticised on the grounds that replication *in vivo* is a non-equilibrium process controlled by stereochemical rather than thermodynamic selection, and should therefore discriminate against non-Watson–Crick base pairs.[37] Some experimental evidence does exist for the greater stacking of BrU compared to uracil (U) in DNA, but the stacking strengths of BrU and T were found to be close.[74, 75] The stacking strength of 5-fluorouracil, whose cytotoxic activity is possibly linked to mispairing with G, is apparently above uracil but below thymine.[74, 76]

An early attempt to combine the notion of enhanced BrU-stacking with a theory of its mutagenicity was made by Sternglanz, Bugg and co-workers.[73, 77, 78] They noticed that the stacking patterns in crystals of pure BrU (and other 5-halouracils) differed from non-halogenated uracils, in that they featured close overlap between the 5-substituent of one molecule and the pyrimidine ring of its stacking neighbour, presumably causing a favourable halogen–aromatic interaction. They proposed a similar motif in the stacking of BrU in DNA, and noted that this would require a slight rotation of BrU relative to its base pairing partner, such that it could not base-pair with adenine due to a steric clash between O4(BrU) and N1(A). Conversely, in this rotated geometry, it could base-pair with guanine through a single hydrogen bond: and the authors proposed *this* as the structure of the BrU-G mispair. However, it is arguable that base pairs require at least two H-bonds to be stable.[35]

The third model – initially proposed, like the rare tautomer hypothesis, in the 1950s[79] – posits an ionised BrU molecule. Here, the base has been deprotonated (formally at N3), and therefore has no H-bond donor sites. In this state, BrU^- can base-pair with G in Watson–Crick geometry (Figure 1.9), but *via* only two hydrogen bonds, both accepted from G: the third H-bond in the rare tautomer mispair is not possible, due to the absence of a proton. The BrU(anion)-G mispair is analogous to the T(anion)-G mispair that may have been observed by Bebenek *et al.*[44] In support of this model, Driggers and Beattie reported that the rate of BrU-G mispairing during chain elongation, catalysed by DNA polymerase, increased as a function of pH when BrU was in the template strand.[80] Similarly, Yu *et al.* found that BrU-G mispairing in phage-derived DNA (using substrate BrU and template G), and in synthetic DNA 26-mers (using template BrU and substrate G), increased with pH.[81] Investigating 5-fluorouracil, correctly base-paired with A in short stretches of synthetic DNA, Kremer *et al.* determined by proton NMR that a significant fraction of FU residues were ionised at physiological pH, and suggested that mispairing of anionic FU with G could play a role in the cytotoxicity of this drug.[82]

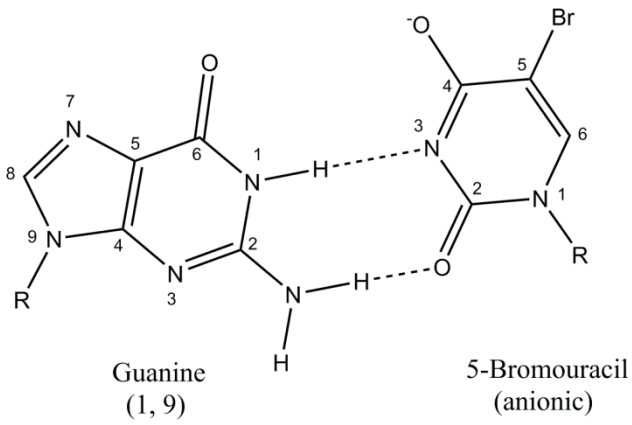


Figure 1.9: Mutagenic base pair between the canonical tautomer of guanine and the deprotonated (anionic) form of 5-bromouracil, in Watson–Crick geometry.

Finally, Sowers *et al.* inferred from proton NMR that the BrU-G mispair in a DNA heptamer was in a pH-dependent equilibrium between wobble and ionised structures, and suggested that both mispairing mechanisms could play a role in mutation.[83]

1.4.1. Note on sequence dependence

It was shown in 1958 that BrU would induce mutations in the DNA of phage particularly often at certain genetic locations, and that these locations were different from the "hot spots" associated with

spontaneous mutations.[53] It was subsequently found that, in eukaryotic DNA, BrU was preferentially incorporated into "moderately repetitive" sequences of nucleotides – those that are repeated between 10^3 and 10^5 times, and are believed to have a regulatory function – and was less easily incorporated into sequences which are more repetitive (appearing around 10^6 times), or which are not repeated at all.[84] Furthermore, Davidson *et al.* have noted that BrU mostly causes AT→GC transitions in prokaryotes, but GC→AT transitions in mammalian cells.[63] A successful theory of BrU mutagenicity should account for the existence of BrU-induced mutational hot spots, and explain why their specific locations within the genome differ from the natural ones.

1.5. Mutagenicity of 5-bromouracil: Theory

In 1998, Orozco *et al.* confirmed by theoretical methods (Hartree–Fock geometry optimisations with MP2 and MP4(SDTQ) single-point energies) that the O4-enol tautomers of 1-methyluracil, 1-methylthymine and 1-methyl-5-bromouracil were all heavily disfavoured relative to the keto tautomers in the gas phase ($\Delta G \approx 11–12$ kcal/mol).[85] Crucially, the free energy penalty of enolisation was only slightly lower for BrU than for T (by about 0.5 kcal/mol), and was not appreciably altered by aqueous solvation (calculated by the self-consistent reaction field method) or the influence of a DNA environment (approximated using Poisson–Boltzmann calculations). They therefore ruled out the involvement of the O4-enol tautomer of BrU in mutagenesis, instead supporting the ionisation model.

In 2004, Hu *et al.* investigated the role of microhydration in the tautomerism of U and BrU.[86] DFT calculations (B3LYP) indicated that the presence of a water molecule between O4 and C5 increased the enolisation penalty of U, compared to the gas phase, thus reducing the equilibrium population of the enol tautomer, but that this "protective" water molecule was electrostatically prevented from occupying the same site in BrU. Conversely, a water molecule bridging O4 and N3 increased the population of the enol tautomer of both species. Hence they concluded in favour of the rare tautomer hypothesis.

In 2005, however, Hanus *et al.* argued that the energetic penalty for hydration at the "protective" water site was so high for both bases that it could be excluded from consideration.[66] On the basis of B3LYP calculations of all the steps involved in the rare tautomeric pathway for mispairing of both BrU and U, accounting for both solvation and the presence of DNA, they supported Orozco *et al.*'s earlier conclusion that tautomerism could not account for the mutagenicity of BrU.

In 2009, Danilov, van Mourik and co-workers[87] investigated the role of bulk hydration of BrU, in an attempt to rationalise the experimental findings[65] of Katritzky and Waring. Arguing that neither microhydration nor continuum solvent methods were sufficient to model the behaviour of nucleic acid bases in bulk water, they fully optimised the keto and enol tautomers of BrU and U embedded in clusters of 50 explicit water molecules. The B3LYP and M05-2X density functionals both determined that, under these conditions, the formation of the enol tautomer was favoured over the keto tautomer for BrU, but not U. This was due to favourable water–water interactions in the hydration shell, which was more densely packed around the enol tautomer of BrU. Thus solvation did not merely increase the equilibrium population of the "rare" tautomer, it actually made it the dominant species. A subsequent study found the same result in clusters of 100 water molecules.[88] No evidence was found for the role of a "protective" water molecule preventing enolisation of U but not BrU, as per Hu *et al.*; instead, the authors argued that the tautomeric shift in favour of enolised BrU under conditions of

bulk solvation was caused by better incorporation of the O4-hydroxyl group into the water network (compared to the O4-carbonyl group), combined with favourable bromine–hydrogen interactions.

This thermodynamic evidence for the stability of BrU* in water was later supplemented by kinetic evidence: molecular dynamics simulations of U and BrU in a periodic water box displayed water-mediated tautomerisation of the latter, but not the former.[89]

In 2013, Nomura *et al.* extended the study of solvent effects on tautomerism from the realm of isolated bases into that of base pairs, of both the wobble and pseudo-Watson–Crick types.[90] The reaction pathway leading from the BrU-G wobble pair to the BrU*-G pseudo-Watson–Crick pair was found to have a lower energetic barrier than the corresponding T-G → T*-G reaction, due to different relative strengths of interaction between the transition state dipole moments and water (modelled as a dielectric continuum). This was taken as further evidence for the rare tautomer hypothesis of BrU-induced mutations.

Clearly, therefore, the hypothesis remains an open possibility. The main part of this thesis will be dedicated to this question. While this began as an experimental subject, computational methods are becoming increasingly important, and may provide insights not available from experiment alone. The next chapter introduces the computational methods used in this thesis.

Chapter 2: Computational methods

2.1. Introduction

To investigate the properties of DNA bases, we have used the tools of computational quantum chemistry. This field consists of the application of computer hardware to solve equations that arise from quantum models of chemical bonding, yielding information about the observable properties of the systems under study.

2.2. Schrödinger equation

The central insight of quantum chemistry is that particles at the microscopic level are fully describable by a single quantity – the wavefunction, Ψ – which encapsulates all the information about their observable properties. This postulate leads to the Schrödinger equation, which specifies the relationship between Ψ and E, the energy of the system – a crucial observable. In its simplest form, the Schrödinger equation is written

$$\hat{\mathrm{H}}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator. This is an eigenvalue equation: by construction, the operation of $\hat{H}$ on the eigenfunction Ψ yields the eigenvalue E. (Other operators can be applied to yield other properties.) The square of the wavefunction, $|\Psi(\mathbf{r})|^2$, can be interpreted as a probability density, being proportional to the probability of finding the particle at a point **r** in space.[91]

In the system's ground state, the wavefunction and energy are labelled Ψ_0 and E_0, but the quantization of energy also results in a set of higher states, Ψ_n and E_n (n = 0, 1, …).

The above formulation holds for a time-independent system in which relativistic effects are ignored. Under these conditions, $\hat{H}$ consists of terms for the kinetic energy (T) and the potential energy (V):

$$\hat{\mathrm{H}}(\mathbf{r}) = T(\mathbf{r}) + V(\mathbf{r}) \tag{2.2}$$

where both terms are a function of the position vector **r**.[92]

For a molecular system containing m nuclei (labelled A) and n electrons (labelled i), $\hat{H}$ is given by

$$\hat{\mathrm{H}} = -\frac{1}{2}\sum_i^n \nabla_i^2 - \frac{1}{2M_A}\sum_A^m \nabla_A^2 - \sum_i^n \sum_A^m \frac{Z_A}{r_{iA}} + \sum_i^n \sum_{j>i}^n \frac{1}{r_{ij}} + \sum_A^m \sum_{B>A}^m \frac{Z_A Z_B}{R_{AB}} \tag{2.3}$$

where M_A are the nuclear masses, Z_A the nuclear charges, r_{iA} the nuclear–electronic distances, r_{ij} the inter-electronic distances, and R_{AB} the inter-nuclear distances. The terms on the right-hand side are, respectively, the electronic kinetic energy, nuclear kinetic energy, nuclear–electronic potential energy, inter-electronic potential energy, and inter-nuclear potential energy. (The Nabla operator for

Cartesian coordinates is defined as $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$, and the equation assumes atomic units: electron rest mass, elementary charge, Bohr radius.)[91]

2.3. Born–Oppenheimer approximation

Since nuclei are considerably heavier than electrons, their motion can be considered static on the timescale of electronic relaxation. This allows the Born–Oppenheimer approximation: the nuclear kinetic energy is zero, and the nuclear–nuclear repulsion is a constant.[91] The resulting simplified form of $\hat{H}$ is the electronic Hamiltonian:

$$\hat{\mathrm{H}}_{\mathrm{electronic}} = -\frac{1}{2}\sum_i^n \nabla_i^2 - \sum_i^n \sum_A^m \frac{Z_A}{r_{iA}} + \sum_i^n \sum_{j>i}^n \frac{1}{r_{ij}} \tag{2.4}$$

which operates on the electronic wavefunction to yield the electronic energy. Given fixed nuclear coordinates, Ψ is now purely a function of the electronic locations and spin states:

$$\Psi_{\mathrm{electronic}} = \Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathrm{n}) \tag{2.5}$$

Here we account for electron spin (an essential consideration for fermions) by replacing the position vectors, **r**, with a set of vectors $\mathbf{x}_i$, which contain both the spatial and the spin (α/β) coordinates of the electrons.

The challenge is to find an appropriate expression for Ψ. At this point, quantum chemical methods diverge into two conceptually different strategies. We use both in this work. The more direct approaches are based on the Hartree–Fock method, which will be discussed first. We will then introduce density functional theory, a rival formulation which treats the electronic density as the central variable.

2.4. Hartree–Fock method

We begin by assuming that the total molecular electronic wavefunction, Ψ, can be decomposed into a set of individual one-electron orbitals, ψ_i. This decomposition takes the form of a Slater determinant:

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathrm{n}) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \ldots & \psi_n(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \ldots & \psi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_n) & \psi_2(\mathbf{x}_n) & \ldots & \psi_n(\mathbf{x}_n) \end{vmatrix} \tag{2.6}$$

As a determinant, this product obeys the Pauli principle of antisymmetry with respect to the interchange of fermions: if two electronic labels (e.g. $\mathbf{x}_1$ and $\mathbf{x}_2$) are swapped, the resulting wavefunction is the negative of the original wavefunction, i.e. Ψ becomes −Ψ.[91]

A crucial approximation in the Hartree–Fock method is that Ψ is described by only *one* Slater determinant.

Given this representation of Ψ, the next requirement is to find optimal terms for ψ_i, the one-electron molecular orbitals (MOs). Here we exploit the variational principle, which states that any trial

wavefunction yields a higher energy than the "true" energy (within the assumptions of the method). From Equation 2.1, the energy, ε, of any trial wavefunction is given by

$$\varepsilon = \frac{\int \Psi^*_{\text{trial}}\, \hat{\text{H}}\, \Psi_{\text{trial}}\, \mathrm{d}\mathbf{x}}{\int \Psi^*_{\text{trial}}\ \Psi_{\text{trial}}\, \mathrm{d}\mathbf{x}} \tag{2.7}$$

By the variational principle, $\varepsilon = E_0$ when $\Psi_{\text{trial}} = \Psi_0$, while $\varepsilon > E_0$ for any other Ψ_{trial}. Hence the optimal molecular orbitals in Ψ_0 are simply those which yield the lowest ε.[91]

We construct the individual MOs as linear combinations of atomic orbitals (LCAOs):

$$\psi_i = \sum_{\mu}^{K} c_{\mu i}\, \varphi_{\mu} \qquad i = 1, 2, \ldots K \tag{2.8}$$

These pre-defined atomic orbitals, φ_μ, are atom-centred, one-electron functions, known as basis functions, which are typically based on hydrogenic orbitals (s, p, d…). The precise form of each ψ_i then depends on the coefficients, $c_{\mu i}$, which modulate the relative contributions of each basis function to each MO.[91]

The coefficients must now be optimised. We first postulate an initial (guess) set of coefficients, in order to generate a trial wavefunction ψ_i. This wavefunction is then introduced into the Hartree–Fock (HF) equations, which are analogous to the Schrödinger equation:

$$\hat{f}_i(\mathbf{x}_i)\psi_i(\mathbf{x}_i) = \varepsilon_i\psi_i(\mathbf{x}_i) \tag{2.9}$$

Here we employ the Fock operator, $\hat{f}_i$, an approximation to the true Hamiltonian operator, given by

$$\hat{f}_i(\mathbf{x}_i) = -\frac{1}{2}\nabla_i^2 - \sum_A^m \frac{Z_A}{r_{iA}} + v^{\text{HF}}(\mathbf{x}_i) \tag{2.10}$$

The first two terms cover the one-electron kinetic and potential energies, respectively. The corresponding one-electron integrals, H_i, are explicitly written as

$$H_i = \left(-\frac{1}{2}\right)\int \psi_i^*(\mathbf{r}_1)\nabla_i^2\psi_i(\mathbf{r}_1)\,\mathrm{d}\mathbf{r}_1 + \sum_A^m(-Z_A)\int \frac{|\psi_i(r_1)|^2}{|\mathbf{r}_1-\mathbf{R}_A|}\,\mathrm{d}\mathbf{r}_1 \tag{2.11}$$

The third term in the Fock operator, the electron–electron repulsion, is given by $v^{\text{HF}}(\mathbf{x}_i)$, the Hartree–Fock potential. This is calculated from the *average* electronic field seen by the *i*th electron due to the presence of all the other electrons. The resulting "mean field approximation" is one of the major approximations in Hartree–Fock theory.[91]

Implicit in $v^{\text{HF}}(\mathbf{x}_i)$ are two types of inter-electronic interaction: Coulombic and exchange. Each is described by a set of two-electron integrals. The Coulomb integrals describe the electrostatic repulsions between electronic charge clouds, and are always positive. They are given by

$$J_{ij} = \iint \frac{|\psi_i(\mathbf{r}_1)|^2|\psi_j(\mathbf{r}_2)|^2}{|\mathbf{r}_1-\mathbf{r}_2|}\,\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \tag{2.12}$$

The exchange integrals describe the effect of switching electronic labels, and arise from the determinantal form of the wavefunction (cf. Equation 2.6):

$$K_{ij} = \iint \frac{\psi_i^*(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)\psi_i(\mathbf{r}_2)\psi_j(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{2.13}$$

For same-spin electrons, the exchange integrals are negative, while for opposite-spin electrons, they are zero. The exchange interaction therefore partially offsets the positive (repulsive) Coulombic contribution to the inter-electronic potential.[91]

The Fock operator for the *i*th electron is now:

$$\hat{f}_i = H_i + \sum_{j=1}^{N} (2J_{ij} - K_{ij}) \tag{2.14}$$

where $N = \frac{n}{2}$ is the number of doubly-occupied MOs, assuming a closed-shell system.

Using linear combinations of one-electron basis functions to express $\psi_i(\mathbf{x})_i$ in the HF equations (c.f. Equation 2.8) yields the Roothaan–Hall (RH) equations:[93, 94]

$$\sum_{\nu}^{K} F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_{\nu}^{K} S_{\mu\nu} c_{\nu i} \qquad i = 1, 2, \ldots K \tag{2.15}$$

where K is the number of basis functions. The RH equations are solved in matrix form: the Fock matrix elements, $F_{\mu\nu}$, are given by

$$F_{\mu\nu} = \int \varphi_\mu^*(\mathbf{r}_1) \, \hat{f} \, \varphi_\nu(\mathbf{r}_1) \mathrm{d}\mathbf{r}_1 \tag{2.16}$$

while the overlap matrix elements, $S_{\mu\nu}$, are given by

$$S_{\mu\nu} = \int \varphi_\mu^*(\mathbf{r}_1) \, \varphi_\nu(\mathbf{r}_1) \mathrm{d}\mathbf{r}_1 \tag{2.17}$$

The usual iterative method for solving the Roothaan–Hall equations is as follows. The initial guess for the orbital coefficients, $c_{\nu i}$, allows the Fock matrix to be constructed. Diagonalising this matrix yields a new set of MO coefficients, $c_{\nu i,}$ by solution of Equation 2.15. These coefficients are then used to construct a new Fock matrix, which yields new values of $c_{\nu i}$. This process is repeated until convergence is reached, i.e. the change in $c_{\nu i}$ from one cycle to the next becomes arbitrarily small.

The resulting solutions to the RH equations are termed a self-consistent field (SCF). The converged coefficients define the HF molecular orbitals, and they yield the electronic energies, ε_i, *via* Equation 2.15. Since each individual ε_i includes the interaction of the *i*th electron with all the others, the sum of ε_i counts the inter-electronic terms twice, so this double-counting must be subtracted from the total electronic energy, E_{el}. The Hartree–Fock energy is then the sum of the total electronic energy and the inter-nuclear potential:

$$E_{\mathrm{HF}} = E_{\mathrm{el}} + \sum_A^m \sum_{B>A}^m \frac{Z_A Z_B}{R_{AB}} \tag{2.18}$$

The Coulomb and exchange integrals are known as four-index integrals, as they can involve up to four different basis functions. The number of basis functions, K, increases linearly with molecular size.

Therefore, the total number of integrals to be solved – and hence the computational expense of a Hartree–Fock calculation – formally scales as the fourth power of K.[91]

2.5. Post-Hartree–Fock methods

A major approximation in HF theory is the neglect of electron correlation. Because the Coulombic potential felt by each electron is treated as a spatially homogenous mean field of all the others, the electron positions are independent of one another, except with respect to exchange. In real systems, however, the electron positions are correlated so as to minimise repulsion, thus lowering the energy. The difference between the "true" and HF energies is the correlation energy:[92]

$$E_{\text{true}} = E_{\text{HF}} + E_{\text{C}} \tag{2.19}$$

where E_{HF} attains its limiting value with an infinite set of basis functions, and the correlation energy E_{C} is negative or zero. Several techniques, termed post-Hartree–Fock methods, have been developed to overcome this weakness of HF theory. They share the common strategy of recovering correlation through the excitations of electrons from the reference (ground-state) HF configuration into unoccupied (virtual) orbitals of higher energy, i.e. lifting the restriction imposed by using a single Slater determinant.

2.5.1. Møller–Plesset perturbation theory

Møller–Plesset perturbation theory calculates the effect of electron correlation as a perturbation to the Fock operators:[92, 95]

$$\hat{\text{H}} = \hat{\text{H}}_0 + \lambda V \tag{2.20}$$

where $\hat{H}_0$ is the sum of one-electron Fock operators, and V is the perturbation, modulated by a quantity λ having a value between 0 and 1. If V is expressed as a Taylor expansion, the resulting energy expression can be truncated at arbitrary order:

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots \tag{2.21}$$

The first two terms on the right-hand side (zeroth and first order) sum to E_{HF} itself. Perturbational corrections to E_{HF} arise from second and higher orders. The choice of truncation order is then denoted MPx, where $x \geq 2$. The most common implementation is MP2, since the second-order correction is the leading term in E_{C}, and higher orders do not always converge.[92]

The MP2 correlation energy is then a sum over doubly excited determinants,[ii] representing the promotion of two electrons from occupied orbitals, i and j, to unoccupied orbitals, a and b:

[ii] Singly excited determinants offer no improvement to the HF energy.

$$E_{\mathrm{C}}^{\mathrm{MP2}} = E^{(2)} = \sum_{i<j}^{occupied} \sum_{a<b}^{virtual} \frac{\int |\psi_i^*(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} [\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) - \psi_b(\mathbf{r}_1)\psi_a(\mathbf{r}_2)] \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{2.22}$$

and the MP2 total energy is simply

$$E_{\mathrm{MP2}} = E_{\mathrm{HF}} + E_{\mathrm{C}}^{\mathrm{MP2}} \tag{2.23}$$

The scaling of MP2 computational expense with respect to the number of basis functions is roughly the fifth power of K.[91]

2.5.1.1. Spin-component scaling

The MP2 correlation energy can be partitioned into contributions from same-spin (SS) and opposite-spin (OS) electron pairs:

$$E_{\mathrm{C}}^{\mathrm{MP2}} = E_{\mathrm{C}}^{\mathrm{SS}} + E_{\mathrm{C}}^{\mathrm{OS}} \tag{2.24}$$

Since Hartree–Fock already accounts for exchange – a purely SS effect – but contains no OS correlation, the OS contribution effectively has further to "catch up" during the perturbative procedure. Truncation at second order preserves this imbalance. It has been found that the errors in $E_{\mathrm{C}}^{\mathrm{MP2}}$ can be reduced by simply scaling the two contributions differently:

$$E_{\mathrm{C}}^{\mathrm{SCS-MP2}} = p_{\mathrm{SS}} E_{\mathrm{C}}^{\mathrm{SS}} + p_{\mathrm{OS}} E_{\mathrm{C}}^{\mathrm{OS}} \tag{2.25}$$

where $p_{\mathrm{SS}} < p_{\mathrm{OS}}$, and typical values are 1/3 and 6/5 respectively. This correction is denoted spin-component scaling (SCS).[96]

2.5.2. Coupled cluster

The coupled cluster method describes the "exact" (fully correlated) wavefunction as a product of the HF wavefunction and an exponential term:[91, 97, 98]

$$\Psi_{\mathrm{CC}} = e^{\hat{\mathrm{T}}} \Psi_{\mathrm{HF}} \tag{2.26}$$

The exponential is in $\hat{\mathrm{T}}$, the cluster operator, which is defined as

$$\hat{\mathrm{T}} = \hat{\mathrm{T}}_1 + \hat{\mathrm{T}}_2 + \hat{\mathrm{T}}_3 + \cdots \hat{\mathrm{T}}_n \tag{2.27}$$

for an n-electron system. The individual excitation operators, $\hat{\mathrm{T}}_i$, generate the Slater determinants required to describe all possible i-tuple excitations from the HF reference. For example, $\hat{\mathrm{T}}_2$ generates all double excitations. Like Møller–Plesset perturbation theory, coupled cluster can be truncated at any arbitrary order up to n. However, being an exponential, the correction to Ψ_{HF} takes the form of a Taylor expansion:

$$e^{\hat{\mathrm{T}}} = 1 + \hat{\mathrm{T}} + \frac{1}{2!}\hat{\mathrm{T}}^2 + \frac{1}{3!}\hat{\mathrm{T}}^3 + \cdots \tag{2.28}$$

For example, in the CCSD truncation,[98] single and double excitations are included:

$$\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2 \tag{2.29}$$

hence

$$e^{\hat{\mathbf{T}}} = 1 + (\hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2) + \frac{1}{2!}(\hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2)^2 + \frac{1}{3!}(\hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2)^3 + \cdots \tag{2.30}$$

Here, the "clustering" of excited configurations results in not only sums, but also products of the excitation operators. These products give rise to higher-level excitations: for example, the square of $\hat{\mathbf{T}}_2$ (third term on the right-hand side) creates quadruple excitations. This method therefore captures a larger fraction of the correlation energy than MP2. However, it does so at greater expense: CCSD scales roughly as the sixth power of K.[91]

The inclusion of operators beyond CCSD is too computationally expensive for most systems. However, a version of coupled cluster which includes a perturbative estimation of the triples term, termed CCSD(T),[99] scales as the seventh power of K, and can be solved for reasonably large systems. This method has come to be known as the "gold standard" of quantum chemistry, due to its high accuracy and computational tractability.[91, 100]

2.6. Density functional theory

Density functional theory (DFT) is an alternative approach to calculating the properties of a system from the nuclear coordinates and electron count. Motivated by the difficulty of working with Hamiltonians, the aim of DFT is to directly establish a link between the energy of the system and the spatially varying *electron density*. This latter quantity, denoted $\rho(\boldsymbol{r})$, represents the total electron probability density at each point r in space. It is arguably a more chemically intuitive concept than the wavefunction, and this is one of the strengths of DFT.[91]

2.6.1. Hohenberg–Kohn theorems

A molecular system can be considered as a set of electrons, interacting both with each other and with an external potential, ν, generated by the fixed nuclei. If the inter-electron interaction and the external potential are both known, the energy can be calculated directly, without recourse to the Hamiltonian that defines the external potential.

The foundation of DFT is two theorems by Hohenberg and Kohn, which prove important points about the relationship between the external potential and the electron density.[101]

The first Hohenberg–Kohn theorem states that any given external potential uniquely determines the electron density. This can be proven by *reductio ad absurdum*: the assumption that two different external potentials give rise to the same electron density is shown to imply a logical contradiction (see Appendix A). This is an existence theorem: given an external potential, *some (unknown) relation* exists to determine the corresponding density. It amounts to the statement that the nuclear coordinates (and their associated charges) fully determine the distribution of the electrons.

The second Hohenberg–Kohn theorem states that the energy associated with any candidate electron density is equal to or greater than the true energy. This is a variational theorem, which follows

logically from the corresponding principle in Hartree–Fock theory: the density is determined by the potential, which is determined by the Hamiltonian – and since that is variational, the density must be too.

The challenge in DFT is therefore to variationally minimise the energy as a function of the electron density, without having to calculate the wavefunction. Since the density itself is a function of the spatial coordinates, ***r***, the energy is a function of a function, termed a functional.[91]

2.6.2. Kohn–Sham method

In the Kohn–Sham (KS) method, the system of interest is considered relative to a fictitious reference system of classical, *non-interacting* electrons, which has the same electron density as the real system.[102] This effectively partitions the energy into a large non-interacting part, and a significantly smaller part arising from quantum electronic interactions. The energy of the non-interacting system is given by

$$E_{\mathrm{ni}}[\rho] = T_{\mathrm{ni}}[\rho] + V_{\mathrm{ne}}[\rho] + J_{\mathrm{ee}}[\rho] \tag{2.31}$$

The terms on the right-hand side are, respectively, the kinetic energy of the non-interacting electrons, the nuclear–electronic interaction (i.e. the external potential), and the classical electron–electron repulsion.[92]

In the real system, the quantum electronic interactions modify both the kinetic energy and (via exchange and correlation) the potential energy. The energy of the real system is thus

$$E_{\mathrm{KS}}[\rho] = E_{\mathrm{ni}}[\rho] + E_{\mathrm{xc}}[\rho] \tag{2.32}$$

where the exchange-correlation energy, $E_{\mathrm{xc}}[\rho]$, contains kinetic and potential contributions:

$$E_{\mathrm{xc}}[\rho] = \Delta T[\rho] + \Delta V_{\mathrm{ee}}[\rho] \tag{2.33}$$

The variational minimisation of $E_{\mathrm{KS}}[\rho]$ then proceeds by the introduction of one-electron orbitals, ψ_i, into a set of pseudoeigenvalue problems, the Kohn–Sham equations:

$$h_i^{\mathrm{KS}}\psi_i(\mathbf{r}_i) = \varepsilon_i\psi_i(\mathbf{r}_i) \tag{2.34}$$

The Kohn–Sham operator is analogous to the Fock operator in HF:

$$h_i^{\mathrm{KS}} = -\frac{1}{2}\nabla_i^2 - \sum_A^m \frac{Z_A}{r_{iA}} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}' + V_{\mathrm{xc}} \tag{2.35}$$

The first three terms on the right-hand side correspond to $T_{\mathrm{ni}}[\rho]$, $V_{\mathrm{ne}}[\rho]$and $J_{\mathrm{ee}}[\rho]$, respectively. The fourth term, V_{xc}, which is analogous to the Hartree–Fock potential, is a derivative of the density:

$$V_{\mathrm{xc}} = \frac{\delta E_{\mathrm{xc}}}{\delta\rho} \tag{2.36}$$

The one-electron orbitals are expressed with a set of basis functions, as in HF theory. The Kohn–Sham operator then acts upon them to generate the energy and the new orbital coefficients. This

operator takes the electron density as an argument (*via* Equation 2.35), so an initial guess for the density must be supplied. The solution of the Kohn–Sham equations then returns a new density, since ρ is a product of the resulting orbitals:

$$\rho(\mathbf{r}) = \sum_{i=1}^{n} |\psi_i(\mathbf{r})|^2 \tag{2.37}$$

The optimal value of ρ can therefore be discovered iteratively, in an analogous manner to the optimisation of molecular orbital coefficients in HF theory.[91, 102]

We have not yet said anything about the exchange-correlation term, V_{xc}. If its form were known exactly, the resulting self-consistent solution to the Kohn–Sham equations would be exact. However, V_{xc} is not known – this is why the KS method begins by partitioning it from the more intuitive terms – and it is presumed unknowable. In its place, many attempts to construct a physically reasonable and computationally tractable approximation to the exchange-correlation functional have been developed.[91]

2.6.3. Local density approximation

The simplest approach is to assume that the exchange-correlation energy at any given point is purely a function of the electron density *at that point*, i.e. the local value of ρ. This is the local density approximation (LDA).[103, 104] The usual form of V_{xc} is then based on the uniform electron gas, a hypothetical phase in which an infinite number of electrons, with some density ρ, moves through a uniformly distributed positive charge. Analytical terms for the resulting exchange energy were derived as long ago as 1927,[105] but the correlation energy must be solved numerically, and involves fitting to empirical constants. Given the means to calculate E_{xc}, we assume that at each point **r** in space it has the value that it would have in a uniform electron gas whose density is also that of point **r**.[91]

2.6.4. Generalised gradient approximation

The LDA can be enhanced by a correction which takes into account the local *gradient* of the density ($\Delta\rho(\mathbf{r})$, the first derivative with respect to **r**). This is the generalised gradient approximation (GGA). A large number of GGA corrections to both the exchange and the correlation energy have been developed, and many are widely used. Like the LDA, these terms tend to have complicated forms. GGA exchange functionals employed in this thesis include Becke88 (B),[106] Becke97 (B97),[107] modified Perdew–Wang (mPW)[108] and Perdew–Burke–Ernzerhof (PBE),[109] while we have used the GGA correlation functional Lee–Yang–Parr (LYP).[110]

The GGA can be enhanced by including terms that depend on either the Laplacian (second derivative) of the density, or the kinetic energy density. This latter quantity, τ, is given by

$$\tau(\mathbf{r}) = \sum_{i}^{\text{occupied}} \frac{1}{2} |\nabla\psi_i(\mathbf{r})|^2 \tag{2.38}$$

Whichever term is used (Laplacian or density), the result is called a meta-GGA functional.[91] In this work, we have used several meta-GGA functionals (actually hybrid-meta-GGA; see below) based on the M05[111] and M06[112] models. These functionals include a dependence on the kinetic energy density

that is heavily parameterised to return accurate results for a large training set of thermochemical and kinetic properties.

2.6.5. Hybrid functionals

Due to the limitations of the uniform electron gas model, the exchange term in DFT, unlike in Hartree–Fock, is not exact. This is a major drawback of both GGA and meta-GGA functionals, and it cannot be remedied simply by total replacement of the exchange functional with the solution of the corresponding HF exchange integrals.[91] However, accuracy can be enhanced by mixing the DFT exchange with a *fraction* of the HF exchange:

$$E_{\mathrm{xc}}^{\mathrm{hybrid}} = (1 - a)E_{\mathrm{xc}}^{\mathrm{DFT}} + aE_{\mathrm{x}}^{\mathrm{HF}} \tag{2.39}$$

where the parameter a can be either empirically optimised or chosen on the basis of theoretical considerations. A functional of this kind is termed a hybrid functional.[91] One of the most popular hybrid-GGA functionals, B3LYP,[113] augments the local density approximation with HF exchange, the Becke88 GGA exchange correction, and LYP correlation, according to three parameters:

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = (1 - a)E_{\mathrm{x}}^{\mathrm{LDA}} + aE_{\mathrm{x}}^{\mathrm{HF}} + b\Delta E_{\mathrm{x}}^{\mathrm{B}} + (1 - c)E_{\mathrm{c}}^{\mathrm{LDA}} + cE_{\mathrm{c}}^{\mathrm{LYP}} \tag{2.40}$$

Correlation is likewise not exact in DFT, and hybrid functionals can be further improved by including a perturbative contribution to the correlation energy.[96] This term is calculated in an analogous manner to MP2 correlation (cf. Equation 2.22), except that the integrals describing the double excitations are evaluated over the Kohn–Sham molecular orbitals. An empirically optimised fraction, a_{c}, of this "PT2" correlation energy is then mixed in with the DFT correlation:

$$E_{\mathrm{xc}}^{\mathrm{DHDF}} = (1 - a_{\mathrm{x}})E_{\mathrm{x}}^{\mathrm{DFT}} + a_{\mathrm{x}}E_{\mathrm{x}}^{\mathrm{HF}} + (1 - a_{\mathrm{c}})E_{\mathrm{c}}^{\mathrm{DFT}} + a_{\mathrm{c}}E_{\mathrm{c}}^{\mathrm{PT2}} \tag{2.41}$$

The result is a double hybrid density functional (DHDF). In this work, we have used the double hybrid mPW2PLYP,[114] which combines mPW and HF exchange with LYP and PT2 correlation.

The successive gain in accuracy upon passing from the LDA to the GGA, then adding meta-GGA terms, then hybrid exchange terms, and finally introducing double hybrids, is often metaphorically likened to climbing rungs on "Jacob's ladder": true chemical accuracy is approached, but at increasing computational cost.[115, 116]

2.6.6. Dispersion corrections

A particular weakness of DFT is the treatment of London dispersion. This is an attractive force, usually attributed to the interaction of instantaneous dipoles arising from fluctuations in electron density (hence it is also described as an induced dipole–induced dipole force). It is a long-range electron correlation effect, and is largely neglected by most GGA correlation functionals.[117, 118] These methods therefore fail to describe complexes that are mostly or entirely bound by dispersion, e.g. noble gas dimers[119] and π-stacked aromatics.[120] The most popular remedy for this takes the form of an empirical, atom-pairwise dispersion correction, resembling the attractive part of the Lennard-Jones potential:[96, 117, 118]

$$E_{\text{disp}} = -S_6 \sum_i^N \sum_j^N \frac{C_6^{ij}}{R_{ij}^6} f_{\text{damp}}(R_{ij}) \tag{2.42}$$

where R_{ij} are the internuclear distances. The scaling factor, S_6, is fitted to experimental van der Waals binding energies (but this factor is neglected in some schemes). The interatomic dispersion coefficients, C_6^{ij}, are calculated from the corresponding atomic terms, typically by

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \tag{2.43}$$

where the atomic terms are computed from the ionisation potentials and polarisabilities. The damping function, f_{damp}, smoothly switches off the attraction at very short ranges. The dispersion correction is then simply added to the DFT exchange-correlation energy in post-processing:

$$E^{\text{DFT-D}} = E^{\text{DFT}} + E_{\text{disp}} \tag{2.44}$$

2.7. Basis sets

The set of atomic basis functions (φ) used to generate the molecular orbitals (ψ), in either a Hartree–Fock or a DFT calculation, is called the basis set. Several methods exist to systematically build up basis sets, whereby increasing the size affords greater accuracy, but comes at a higher computational cost.

2.7.1. Atom-centred basis sets

In their form, atom-centred basis functions resemble the hydrogenic orbitals (s, p, d …) yielded by the exact solutions to the one-electron Schrödinger equation. An ideal set of functions to represent these would be Slater-type orbitals (STOs), which have the appropriate radial decay in $e^{-\mathbf{r}}$. However, the basis functions used in Hartree–Fock and DFT calculations are usually constructed from linear combinations of Gaussian-type orbitals (GTOs). A three-dimensional GTO, with the distance from the origin given in Cartesian coordinates (**x**, **y**, **z**), has the general form

$$\varphi(\mathbf{x},\mathbf{y},\mathbf{z};\ \alpha,i\,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k}\, i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2} \mathbf{x}^i\mathbf{y}^j\mathbf{z}^k e^{-\alpha(\mathbf{x}^2+\mathbf{y}^2+\mathbf{z}^2)} \tag{2.45}$$

where α controls the width of the function, and *i*, *j* and *k* are non-negative integers. When *i*, *j* and *k* are all zero, the GTO is effectively an s-orbital; when they sum to one, they create a p-orbital; when they sum to two, they create a d-orbital, etc. These functions have the advantage that the product of two GTOs is itself a GTO. Although individual GTOs ("primitives") decay too fast, as $e^{-\mathbf{r}^2}$, their linear combinations ("contracted GTOs") can be optimised to resemble well-behaved STOs.[91]

In *minimal* basis sets, only one contracted GTO is provided for each orbital (1s, 2s, 2p…). However, in this work we have used *split-valence* basis sets, in which the valence (and in some cases also the core) orbitals are described by two or more contracted GTOs. This allows extra flexibility, improving the valence orbital behaviour in chemical bonding environments. Basis sets in which the valence orbitals consist of two, three, etc. contracted GTOs are known as double-ζ, triple-ζ basis sets, etc.[91]

In Pople-type basis sets, the core orbitals are described by single functions, while the valence orbitals are described by two or more. These basis sets are indicated by acronyms of the form "6-31G". In that (double-ζ) example, the core orbitals are single functions contracted from six GTOs, while the valence orbitals are described by one contraction of three GTOs and one single GTO. An example of a Pople-type triple-ζ basis set is 6-311G.[91]

Dunning-type correlation-consistent basis sets contain, in addition to a basic set of functions optimised for the HF energy, an extra set of functions that make a similar contribution to the correlation energy. These functions improve the correlation in a systematic, hierarchical way as a function of ζ. These basis sets are denoted "cc-pV*n*Z", where *n* = D, T, Q…, for double-ζ, triple-ζ, etc. For example, cc-pVDZ has two s- and two p-functions and one d-function, while cc-pVTZ has three s- and three p-functions, two d-functions and one f-function, and so on.[91, 121]

Flexibility can be enhanced by adding orbitals with an angular momentum quantum number that is one higher than the valence orbitals, e.g. d-functions on first-row elements. These *polarization functions* help to account for the greater degrees of freedom needed by molecular orbitals – which are functions of multiple atomic positions – compared to monatomic orbitals. They are included by default in Dunning-type basis sets. When added to Pople-type basis sets, they are indicated either by "*", e.g. 6-31G*, or more explicitly by quantum number, e.g. 6-31G(d,p), which contains d-functions on first-row elements and p-functions on hydrogen.[91]

The treatment of weakly-bound systems, in which some valence electrons may be localised far from the nuclei, can be improved by adding shallow *diffuse functions*, which have small exponents. In Pople-type basis sets, the sign "+", e.g. 6-31+G(d), indicates the addition of one extra set of diffuse s- and p-functions to non-hydrogenic atoms. In Dunning-type basis sets, diffuse functions are shown by the prefix "aug-", e.g. aug-cc-pVDZ; this incurs a greater computational expense, since one set of diffuse functions is added for each angular momentum already present, including the polarization functions.[91] In this thesis we will abbreviate aug-cc-pV*n*Z as "a*n*Z".

The combination of a method with a basis set is denoted by a single slash (/), e.g. MP2/cc-pVDZ.

2.7.2. Plane wave basis sets

Calculations on periodic systems – e.g. single-point calculations on crystals, or the dynamics of a liquid solution modelled using periodic boundary conditions – often use plane wave basis sets. Here, the molecular orbitals are expanded in a set of sine and cosine functions, extending infinitely throughout space, whose kinetic energy contribution is determined by the frequency of oscillation of the wave:

$$\psi(\mathbf{x}) = A\cos(\mathbf{K}\mathbf{x}) + B\sin(\mathbf{K}\mathbf{x}) \tag{2.46}$$

where **K** is the wave vector and *A* and *B* are optimisable coefficients. **K** is related to the size of the unit cell, *t*, by

$$\mathbf{K}\cdot t = 2\pi m \tag{2.47}$$

where *m* is a positive integer. **K** in turn determines the energy by

$$E = \frac{1}{2}\mathbf{K}^2 \tag{2.48}$$

A maximum-energy cut-off is usually enforced.[92, 122]

2.7.3. Pseudopotentials

When plane waves are used for molecular calculations within a periodic system, the core orbitals are usually represented by a fitted function known as a *pseudopotential* (PP) or *effective core potential* (ECP). The slowly varying character of plane waves means that a very large number of them must be used to model the rapid oscillations of the core wavefunctions (many more than are needed in atom-centred basis sets), which renders them inefficient for this purpose. Since core electrons are usually of little chemical interest, it is expedient to replace them with a mathematically simpler potential, which is fitted to exactly recover the valence behaviour above a certain electron–nucleus cut-off distance.

This fitting requires the initial calculation of an all-electron reference state. The core wavefunctions are subsequently replaced by the PP, while the Hamiltonian for each valence electron now contains an additional potential term accounting for the core. Hence the valence shell experiences an "effective core", containing not only the nucleus itself but also the inner electrons.[91]

In this work, we have used *norm-conserving* pseudopotentials, which are constructed such that the norm (roughly speaking the square) of the PP matches that of the reference wavefunction both inside and outside the cut-off radius.

2.7.4. Basis set superposition error

In the supermolecular method, the interaction energy, ΔE, of a dimer, A/B, is defined as the energetic difference between the dimer and the sum of the infinitely-separated monomers, A and B:

$$\Delta E = E_{\mathrm{A/B}} - E_{\mathrm{A}} - E_{\mathrm{B}} \tag{2.49}$$

Unfortunately, any calculation that uses a finite, atom-centred basis set will be contaminated by basis set superposition error (BSSE).[91] The basis functions of A and B will overlap in the dimer, allowing each monomer to use those of the other. This effectively enlarges the basis set used to calculate $E_{\mathrm{A/B}}$, causing a mismatch between the dimer and monomer calculations. By the variational principle, any improvement to an energy calculation lowers the energy, so the BSSE causes an artificial attraction between the monomers.

This becomes clearer if the uncorrected energy in Equation 2.49 is written more explicitly as

$$\Delta E^{\mathrm{uncorrected}} = E_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B}) \tag{2.50}$$

The superscripts indicate the basis set used for each calculation: the dimer is treated with the enlarged basis set, {A/B}, while the monomers are treated with their own basis sets, {A} and {B}. This mismatch is responsible for the BSSE.

The parentheses indicate the geometries at which the energies are calculated: the dimer is in the dimer-optimised geometry, (A/B), while the monomers are in their geometries optimised at infinite separation, (A) and (B) (see Section 2.10 for geometry optimisation).

The usual remedy for the BSSE is the counterpoise procedure (CP).[123] Each monomer energy is re-calculated using the enlarged basis set, through the inclusion of "ghost orbitals": basis functions with no associated atoms, placed in locations corresponding to the other monomer in the dimer geometry.

The monomer terms using the enlarged basis set must be calculated at the dimer-optimised geometries, in order to include the ghost orbitals. These terms are denoted $E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B})$ and $E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B})$. However, the interaction energy is defined relative to the monomers in their own optimised geometries. We therefore need to explicitly introduce the deformation energy, which was implicit in Equations 2.49 and 2.50. This is the unfavourable term arising from the geometric change that each monomer undergoes when passing from infinite separation to the dimer. It requires two further calculations: the energies of each monomer in their own basis sets at their dimer-optimised geometries, denoted $E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B})$ and $E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B})$. The deformation energy is then

$$E_{\mathrm{def}} = E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B}) + E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B}) \tag{2.51}$$

Hence the counterpoise procedure for a dimer requires, in addition to the three terms on the right-hand side of Equation 2.50, four extra calculations: the two monomer energies with their own basis sets, and the two monomer energies with the enlarged basis set, all calculated at the dimer-optimised monomer geometries.

The final counterpoise-corrected interaction energy, with the BSSE removed and the deformation energy included, is then

$$\Delta E^{\mathrm{CP}} = E_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) + E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B}) + E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B}) \tag{2.52}$$

where the last four terms account for deformation (cf. Equation 2.51). The BSSE, meanwhile, is given by

$$\mathrm{BSSE} = E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) + E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B}) \tag{2.53}$$

which is the negative of the terms between second and fifth on the right-hand side of Equation 2.52. By subtracting the right-hand side of Equation 2.53 from that of Equation 2.50, it can be seen that the counterpoise-corrected interaction energy in Equation 2.52 is equivalent to subtracting the BSSE from the uncorrected energy:

$$\Delta E^{\mathrm{CP}} = \Delta E^{\mathrm{uncorrected}} - \mathrm{BSSE} \tag{2.54}$$

The basis functions of different atoms can also overlap within a single molecule, and because this overlap is conformation-dependent, the resulting intramolecular BSSE can distort conformational analysis.[124, 125] However, no general prescription for this problem has found wide acceptance.

2.8. Resolution of the identity (density fitting)

Resolution of the identity (RI) is a family of techniques for reducing the expense of HF-based calculations (e.g. MP2) as a function of basis set size.[126] The key step is the factorisation ("resolution") of the four-index-two-electron integrals into a sum over products of three-index overlap

integrals and three-index-two-electron integrals. The one-electron charge densities from the computationally intensive four-index integrals are now expanded in an auxiliary ("fitting") basis set within the less demanding three-index integrals. RI-MP2 is therefore somewhat more efficient than MP2. However, because the scaling of the exchange calculations is not formally reduced, the method still scales with the fifth power of K, hence the computational savings are usually modest. The fitting basis set is typically several times larger than the atomic orbital basis set; nevertheless, its incompleteness represents a (small) source of error.[127] Since the essential challenge of RI is to minimise this error, the method is alternatively referred to as "density fitting" (prefixed "df-").[128]

2.9. Local orbital approximation

Orbital localisation is a method for reducing the computational expense of post-Hartree–Fock calculations. In the procedure of Schütz *et al.*, based on the work of Pulay, the occupied MOs from a converged SCF calculation are *localised* before the treatment of electron correlation begins. This is done by projecting out the occupied space in the atomic orbitals, generating a new set of projected AOs.[127, 129] Two simplifications in the treatment of correlation between electron pairs, ij, can now be made. Firstly, correlation is calculated using only a subset ("domain") of the virtual MOs. Secondly, a hierarchical treatment of the electron pairs, classified according to their inter-electronic distance, r_{ij}, is employed: close ("strong") pairs are treated at the highest level of theory (e.g. CCSD); distant ("weak") pairs at a lower level (e.g. MP2); and very distant pairs are neglected. Since the number of very distant pairs grows with system size much faster than the others, the scaling of expense as a function of system size is sharply reduced by this step (also for the exchange integrals). Orbital localisation also reduces the BSSE.[127, 129]

Local methods are prefixed by "L", e.g. LMP2[130] or LCCSD.[131] A local treatment of perturbative triples is also possible, in which only the excitations of spatially proximate triplets of electrons are considered, and they are restricted to localised domains. This leads to the LCCSD(T) method.[132] The auxiliary basis set in a density fitting calculation can also be localised, leading to "local density fitting" methods, such as df-LMP2,[128] which is more efficient than either df-MP2 or LMP2.

2.10. Geometry optimisation

It is often desirable to find the energy minimum of a system – the geometry with the lowest energy – as this is expected to be the most populated state at thermodynamic equilibrium. The search for this structure is known as geometry optimisation. Having calculated, at some level of theory, the energy E of an initial geometry k, the path towards the minimum is followed by calculating the first and (usually) second derivatives of E with respect to geometric displacements.

If the geometry is defined by a set of internal coordinates, q_i, consisting of bond lengths, angles and dihedrals, there are $n = 3N - 6$ coordinates, where N is the number of atoms. Now, the first derivative of E as a function of the geometry (i.e. the force) is a vector, $\mathbf{g}$, containing all n values of $\partial E/\partial q_i$. This vector is zero if and only if the given geometry is a stationary point on the energy hypersurface. Additionally, we can construct an $n \times n$ matrix, $\mathbf{H}$, known as the Hessian, containing all second derivative terms, $\partial^2 E/\partial q_i \partial q_j$. At a given stationary point, the eigenvalues of $\mathbf{H}$ are all positive if and only if the stationary point is a minimum.[91]

The challenge is therefore to find a geometry for which $\mathbf{g}$ = 0 and the eigenvalues of $\mathbf{H}$ are all positive. Starting from some geometry defined by $\mathbf{q}^{(k)}$, the next step towards the minimum is derived from $\mathbf{H}^{(k)}$ and $\mathbf{g}^{(k)}$ by

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} - (\mathbf{H}^{(k)})^{-1}\,\mathbf{g}^{(k)} \tag{2.55}$$

Each successive $\mathbf{g}$ is, generally speaking, smaller than the last. The optimisation is deemed to have converged when the maximum component and root mean square of both $\mathbf{g}$ and the geometric displacement, $\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}$, each fall acceptably close to zero.

Several algorithms exist for speeding up this search. In the popular Berny algorithm,[133] the next step is taken to be the sum of linear and quadratic steps, where the linear step is obtained by a polynomial fit to the energies of the latest point and the previous point, and the quadratic step is then based on the linear step. When the latest point is not the lowest in energy so far, certain restrictions are applied to the acceptability of the next step. The step sizes are also limited to a maximum value, known as the trust radius.

Full analytic Hessians are very expensive to calculate and invert, so the initial $\mathbf{H}$ is sometimes calculated at a lower level of theory. After the first optimisation step, the updated values of $\mathbf{H}$ are then often estimated, based on the changes in $\mathbf{g}$, rather than being computed exactly.[91]

Geometry optimisation algorithms in general have no way of enforcing convergence to the *global* minimum for any system. When multiple minima exist on the energy hypersurface, an optimisation will usually converge to whichever one happens to be closest to the starting structure. A global search therefore requires multiple optimisations, starting from sufficiently different structures.[91]

When a system is optimised at a certain level of theory, and the energy of the optimised geometry is then determined at a higher level of theory, the composite method is denoted by a double slash (//) with the optimisation method on the right-hand side. For example, MP2/6-31G//HF/6-31G indicates an MP2 single-point energy calculation on an HF geometry.

2.11. Continuum solvent models

For calculations on liquid solutions, full atomistic representation of each solvent molecule (explicit solvation) is usually too expensive at the quantum level. This problem can be tackled by continuum methods: these are implicit solvent models, treating the solvent simply as a dielectric field in which the solute is embedded. In this way, the key feature of the solute/solvent interaction – the mutual polarisation of their electric fields – can be described without recourse to a full solvent wavefunction.[91]

The solute is placed in a cavity within the electric field, the size and shape of which is determined by overlapping spheres surrounding each solute atom. At each grid point on the cavity surface, the local charge density (electric field) of the solute (determined from its gas-phase SCF orbitals) then interacts with that of the solvent, according to the Poisson equation:

$$\nabla\varepsilon(\mathbf{r}) \cdot \nabla\varphi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.56}$$

where ε is the solvent's dielectric constant, φ is the electrostatic potential, and ρ is the charge density of the solute. To minimise φ, the solvent's electric field then relaxes in response to that of the solute.

The extent to which it can be polarised in this way depends on ε. The solute's electric field in turn adjusts to the change in the solvent, and so on. The mutually optimum charge distributions of solute and solvent must therefore be discovered iteratively at each grid point, leading to a self-consistent reaction field (SCRF).[91]

Having solved this electrostatic problem, the solvation energy can then be expressed as the sum of the electrostatic potential and three other terms: one accounting for the creation of the cavity (in itself unfavourable), and two empirical terms accounting for exchange-repulsion effects and dispersion effects between solvent and solute.[92, 134] This is according to:

$$\Delta E_{\text{sol}} = E_{\text{electrostatic}} + E_{\text{cavitation}} + E_{\text{ex-rep}} + E_{\text{disp}} \tag{2.57}$$

One of the most popular SCRF methods is the polarisable continuum model (PCM), also known as MST, in reference to its originators, Miertuš, Scrocco and Tomasi.[91, 134]

2.12. Molecular dynamics

Molecular dynamics (MD) is a method of simulating the behaviour of a chemical system, e.g. a liquid solution, as a function of time. The system is modelled as a set of classically interacting particles, with known mass and acceleration, whose positions change throughout a succession of discrete timesteps following Newtonian principles. Exploring the system's phase space in this way allows the calculation of certain kinetic data which cannot be obtained solely from the ground-state energy minimum, which is a time-independent quantity.

MD is based on Newton's second law, **F**=m**a**, which relates the force acting on a body to the mass of that body and the resulting acceleration. The acceleration, **a**, is a vector, defined as

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} \tag{2.58}$$

where **v** is velocity, *t* is time and **r** is position. If the energy of the system, in a given geometry, at some time *t*, can be calculated (e.g. by DFT), the force acting on each particle can be derived. This can then be used to propagate the motion of the system through subsequent timesteps, by updating **a** (as a function of **F**) and **v** (as a function of **a**) at each new value of *t*.

Various algorithms exist for estimating the position vector, $\mathbf{r}_i$, at each new timestep. Note that since $\mathbf{r}_i$ actually depends not only on **v** and **a** (its first and second derivatives with respect to time), but also on higher derivatives, none of these algorithms are truly exact. One of the most popular is the leapfrog algorithm:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + (\mathbf{v}_{i+\frac{1}{2}} \times \Delta t) \tag{2.59}$$

$$\mathbf{v}_{i+\frac{1}{2}} = \mathbf{v}_{i-\frac{1}{2}} + (\mathbf{a}_i \times \Delta t) \tag{2.60}$$

Here, $\mathbf{r}_i$ and **v** are updated out of phase with each other by half a timestep, i.e. each quantity "leapfrogs" over the other. Since **v** appears explicitly, the leapfrog algorithm is useful for simulations in which the temperature needs to be controlled, since temperature is a function of velocity, *via* kinetic energy.[91]

2.12.1. Car–Parrinello Molecular Dynamics

Car–Parrinello Molecular Dynamics (CPMD) is a technique for solving the equations of motion in an MD simulation, which exploits a quantum chemical treatment of the electronic structure, but without a prohibitively expensive set of SCF calculations. The atomic forces, **F**, are derived from DFT calculations on the complete system, using plane waves for the valence electrons and pseudopotentials for the core. At the first step of the simulation, the Kohn–Sham orbital coefficients, *c*, are fully optimised in the conventional way for the starting geometry, yielding a set of atomic forces to kick-start the dynamics of nuclear motion. However, at subsequent timesteps, the orbitals are *not* self-consistently re-optimised in response to the changing geometries. Rather, the coefficients are assigned fictitious masses, μ, and their optimisation is treated as a dynamical problem, in which the iterative steps are analogous to conventional MD timesteps. They can therefore be described by a quasi-Newtonian law of motion:

$$\mathbf{F} = \mu\ddot{\psi}_\iota = -\frac{\partial E(\{\psi_i\},\{R_I\})}{\partial \psi_i^*} + \sum_j \Lambda_{i,j}\,\psi_j(r) \tag{2.61}$$

where $\ddot{\psi}_\iota$ is the "acceleration" of the orbital coefficients (ψ_i), ∂E are the orbital energies, R_I are the nuclear positions, and $\Lambda_{i,j}$ are undetermined Lagrange multipliers (necessary to maintain orbital orthonormality). In effect, the degrees of freedom of the coefficients are treated as dynamical variables, which evolve over time together with the nuclear positions. These two trajectories are explicitly coupled via an extended Lagrangian, so that the changes in the nuclear positions result in non-zero forces acting on the orbital coefficients. The coefficients now "chase" the electronic ground states, E_0, of the updated geometries. If sufficiently small timesteps and fictitious masses are used, the acceleration of *c* towards E_0 is so efficient that the system's electronic structure stays acceptably close to the "true" (Born–Oppenheimer) ground state, but at a much lower cost than would be needed for full SCF convergence at each timestep.[91, 122, 135]

2.13. Thermochemistry

Standard quantum chemical calculations, of the type outlined in Sections 2.1 – 2.11, implicitly take place at 0 K. At higher temperatures, however, thermal effects can have a significant influence on the relative energies of different species, so that actual chemical behaviour is driven by enthalpy or free energy. The connection between the zero-temperature potential energy surface – as calculated with quantum theory – and the finite-temperature enthalpy and free energy surfaces – as measured by experiment – is made through various thermochemical relationships. In principle, the temperature-dependent energetic terms arise from three kinds of molecular motion: translation, rotation and vibration. Here, we will only consider vibration.

The vibrational corrections can be approximated as follows. For a system occupying a potential energy minimum, the vibrational enthalpy is related to the frequencies of the molecular vibrational modes (which are calculated, within the harmonic approximation, by standard quantum chemical packages) *via*:

$$H_{\text{vib}}(T) = R\sum_i \left(\frac{h\nu_i}{2k} + \frac{h\nu_i}{k}\frac{1}{e^{\frac{h\nu_i}{kT}}-1}\right) \tag{2.62}$$

where T is the temperature, R is the gas constant, h is Planck's constant, v_i are the harmonic frequencies of the vibrational modes, and k is Boltzmann's constant.

The vibrational entropic corrections, $S_{\text{vib}}(T)$, can also be obtained from the frequencies:

$$S_{\text{vib}}(T) = R \sum_i \left[\frac{h\nu_i}{kT} \frac{1}{e^{\frac{h\nu_i}{kT}} - 1} - \ln\left(1 - e^{-\frac{h\nu_i}{kT}}\right) \right] \tag{2.63}$$

The Gibbs free energy (G) is then:

$$G = H_{\text{vib}} - T\, S_{\text{vib}} \tag{2.64}$$

Chapter 3: DFT study of tautomerism in 50-water clusters

3.1. Background

Calculations by Orozco *et al.*[85] and Hanus *et al.*[66] have confirmed that the 4-enol (mutagenic) tautomer of BrU is heavily disfavoured in the gas phase. In the former study, the electronic energy of tautomerisation, ΔE (defined as $E_{\text{enol}} - E_{\text{keto}}$), was 11.5 kcal/mol at the MP4(SDTQ)/6-311+G(d,p)//HF/6-311+G(d,p) level of theory (for 1-methyl-5-BrU). In the latter study, ΔE was found to be 11.54 kcal/mol at the MP2/cc-pVTZ level of theory (for plain BrU, without the 1-methyl substituent).[iii]

To assess the rare tautomer hypothesis of BrU-induced mutagenesis, we need to calculate and compare the extent to which BrU and the natural bases thymine (T) and uracil (U) tautomerise. Although U is not found in DNA, it is an essential component of RNA, the molecule responsible for converting the instructions in the genetic code into the actual synthesis of proteins. T-G and BrU-G mispairs in DNA become U-G mispairs following transcription into RNA. Some theoretical studies of BrU mutagenesis have focused on U rather than T as the natural base to which BrU is compared.

We have calculated tautomerisation energies for the non-methylated forms of BrU and U, optimised in the gas phase at several levels of theory, using Gaussian 09.[136] The calculated energies are shown in Table 3.1.

	$\Delta E^{\text{gas}}_{\text{keto}\rightarrow\text{enol}}$ (BrU)	$\Delta E^{\text{gas}}_{\text{keto}\rightarrow\text{enol}}$ (U)
BLYP	12.3	12.1
B3LYP	12.5	12.2
M05-2X	11.3	10.7
M06-2X	11.4	10.8
M06-L	13.0	12.7
PBE0	12.5	12.0

Table 3.1. Gas-phase enolisation energies of BrU and U in kcal/mol. All functionals employ the 6-31G(d,p) basis set.

At all six levels of theory, the energy penalty for gas-phase enolisation is actually slightly higher for BrU than for U. However, the importance of water in affecting tautomeric ratios has been recognised ever since Katritzky and Waring[65] found evidence for an elevated enol population of 1-methyl-5-BrU in aqueous solution. In the context of water-assisted base tautomerisation, Furmanchuk *et al.* have argued that, although water is probably excluded from the replication machinery, it is present in bulk elsewhere in the cell, and could be involved in pre-replicative base tautomerism.[137] It is known from experiment that when DNA is crystallised under hydrous conditions, water is bound tightly to the DNA surface at a concentration of up to 10 water molecules per nucleotide.[45]

[iii] "Keto" indicates the 2,4-dioxo tautomer; "enol" indicates the 2-oxo-4-hydroxy tautomer with N3 in the imine form.

The theoretical studies by Orozco *et al.* and Hanus *et al.* did also take solvation of BrU into account, but they used methods which did not explicitly treat the interaction between the base and the complete solvation shell which surrounds it. Orozco *et al.* used an SCRF implementation of the polarisable continuum model[138] of water (PCM).[85] On this basis, they calculated that aqueous solvation of 1-methyl-5-BrU only slightly decreased the free energy penalty of tautomerisation, ΔG, from 11.8 kcal/mol in the gas phase to 11.0 kcal/mol in water, at the MP4(SDTQ)/6-311+G(d,p)//HF/6-311+G(d,p) level of theory. The corresponding free energy decrease in water was actually greater for 1-methyluracil (from 11.2 to 9.8 kcal/mol), and very similar for 1-methylthymine (from 12.5 to 11.5 kcal/mol). Together, these results suggest (i) that the presence of either methyl or bromine at the 5-position actually increases the free energy penalty of tautomerisation of 1-methyluracil, (ii) that the penalty is only slightly smaller for 1-methyl-5-BrU than for 1-methylthymine, and (iii) that aqueous solvation does not sensibly change the overwhelming preference for the keto tautomers of all three species.

Similarly, Hanus *et al.* calculated a tautomerisation free energy of 10.01 kcal/mol for non-methylated BrU at the MP2/cc-pVTZ level of theory, combined with the COSMO[139] continuum solvation model.[66] This penalty was smaller than the 11.95 kcal/mol that they had obtained with the same method in the gas phase, but this reduction in free energy was deemed "not decisive". The same study also considered the interaction of a single explicit water molecule with the keto and enol tautomers of BrU and U in the gas phase. For BrU, such an interaction was found to either increase or decrease the electronic energy (not free energy) of tautomerisation, by an amount ranging from +3.8 to −3.0 kcal/mol, depending on the position of the water molecule. This was again not considered to be significant evidence for the formation of the rare tautomer of BrU in water, especially as the hydration site most conducive to tautomerisation was not the most favourable one for the keto tautomer, i.e. the starting state. (For U, all the hydration sites reduced the tautomerisation energy, but by less than for BrU.)

Our own calculated tautomerisation energies for U and BrU, with the inclusion of PCM aqueous solvent, at the same levels of theory that were considered at the start of this chapter, are shown in Table 3.2. All calculations were performed with Gaussian 09.

	$\Delta E^{\mathrm{PCM}}_{\mathrm{keto}\rightarrow\mathrm{enol}}$ (BrU)	$\Delta E^{\mathrm{PCM}}_{\mathrm{keto}\rightarrow\mathrm{enol}}$ (U)
BLYP	12.0	11.3
B3LYP	12.1	11.2
M05-2X	10.8	9.6
M06-2X	10.8	9.6
M06-L	12.6	11.8
PBE0	12.0	11.0

Table 3.2. Aqueous (PCM) enolisation energies of BrU and U in kcal/mol. All functionals employ the 6-31G(d,p) basis set.

At each level of theory, the inclusion of PCM solvent reduces the energy penalty for tautomerisation of each base compared to the gas phase (cf. Table 3.1). However, this effect is always greater for U, so that the population of the enol tautomer is still expected to be larger for U than for BrU, and by an even greater amount than in the gas phase. Therefore, on the basis of Tables 3.1 and 3.2, the rare tautomer model of the BrU-G and U-G mispairs would actually predict a greater rate of mutagenesis for U than for BrU – contrary to fact – assuming that mispairing was the only cause of mutations, and accounting for the hydration of DNA.

However, it must be borne in mind that water-assisted tautomerisation is a chemical reaction involving the making and breaking of bonds, possibly accompanied by stabilisation of the rare tautomer through orbital overlap with the solvent molecules. This idea was explored by Hu *et al*. in a Natural Bond Orbital (NBO) analysis of uracil interacting with an explicit (DFT) water molecule.[140] They calculated that a water molecule located between N3-H3 and C4=O4 (termed "W1") stabilised the O4-enol tautomer relative to the keto tautomer, while a water molecule located between C4=O4 and C5-H5 ("W2") had the opposite effect. This difference was attributed to the different orbital interactions between each tautomer and each water position: W1 took part in stronger $n \rightarrow \sigma^*$ interactions with the enol tautomer, while W2 took part in stronger $n \rightarrow \sigma^*$ and $\sigma \rightarrow \sigma^*$ interactions with the keto tautomer. Thus W1 encouraged tautomerisation, while W2 protected the base against tautomerisation. However, crucially, their subsequent study showed that for BrU, the "protective" W2 position was electrostatically unfavourable for occupation by water – any attempt to optimise a water molecule in that position resulted in its migration elsewhere.[86] The authors therefore concluded that hydration of the bases should increase the equilibrium population of the enol tautomer of BrU, but not that of U, compared to the gas phase, since the keto-stabilising effect of W2 would not apply to BrU. (The authors confusingly swapped the labels of W1 and W2 between the two papers; here we use the labels from the latter paper, Ref.[86])

The findings of Hu *et al.* emphasise that the bulk solvation of nucleic acid bases cannot be fully modelled by continuum solvation methods, as per Orozco *et al*. In continuum methods, the solvent is simply modelled as an electric field perturbing the electronic energy levels of the solute, but molecular orbital interactions between solvent and solute are not captured. Although the inclusion of one or several explicit water molecules around the base (termed microhydration), as per Hu *et al*. and Hanus *et al*., does partially account for orbital overlap, it still does not fully describe bulk solvation (even when combined with a continuum method), because it neglects solvent–solvent chemical interactions within the hydration shell.

Danilov, van Mourik and co-workers subsequently demonstrated that solvation in a cluster of 50 explicit water molecules (termed a nanodroplet) actually reversed the tautomeric preference of BrU, but not U, in favour of the enol, at both the B3LYP/6-31G(d,p) and M05-2X/6-31G(d,p)//B3LYP/6-31G(d,p) levels of theory.[87] With the latter method, the counterpoise-corrected formation energy of the BrU-enol$(H_2O)_{50}$ cluster was *lower* than that of the BrU-keto$(H_2O)_{50}$ cluster by 2.8 kcal/mol. (This study considered electronic rather than free energies, due to the prohibitive cost of vibrational calculations on such large systems.) With both methods, the counterpoise correction was necessary to obtain the reversal of tautomeric preference. Realistic starting structures for the geometry optimisations, stabilised by hydrogen bonds in the water network, were obtained using Monte Carlo simulations containing 400 water molecules, from which the nearest 50 to the centre of the base were chosen at the end of the simulations. The cluster size of 50 waters was justified by an earlier molecular dynamics study showing that the stability of RNA base *pairs* in nanodroplets reached its bulk water value at around 100 waters.[141]

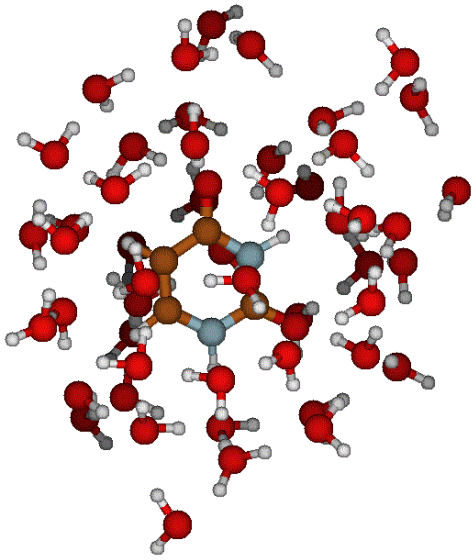


Figure 3.1: U-keto$(H_2O)_{50}$ cluster optimised at the B3LYP/6-31G(d,p) level by van Mourik *et al.*[88]

A more detailed follow-up study by van Mourik *et al.* then considered nanodroplets containing either 50 or 100 waters, still with single U or BrU bases, using the BLYP, B3LYP and M05-2X functionals and 6-31G(d,p) basis set.[88] For both cluster sizes, and at all levels of theory, the earlier trend was confirmed: formation of the enol-containing cluster was favoured for BrU (in contrast to the gas phase), but the keto-containing cluster was favoured for U. The tautomerisation energies, $\Delta E_{\text{keto}\rightarrow\text{enol}}$, for BrU ranged from −14.7 kcal/mol for the 50-water cluster at the BLYP/6-31G(d,p) level of theory to −1.4 kcal/mol for the 100-water cluster at the same level of theory. The tautomerisation energies for U were +1.8 and +15.8 kcal/mol at this level of theory, in 50-water and 100-water clusters, respectively. Decomposition of these energies into various components – deformation of base and water upon complexation, base–water interaction, water–water interaction, and gas-phase tautomerisation energy – showed that the water–water interactions were crucial. Base–water interactions favoured the enol for both U and BrU (in agreement with Hanus *et al.*),[66] while water–water interactions favoured the keto form of U but the enol form of BrU. For the stability of enolised BrU, base–water interactions were significantly more important in the larger cluster, while water–water interactions were important in both clusters. From visual inspection, the authors inferred that enolisation of BrU in the 50-water clusters was favourable due to a cooperative water chain around the 4-hydroxyl group, which acts like a water molecule. In the 100-water clusters, the enhanced stability of the enol was attributed to the interaction between the Br atom and the H atoms of two water molecules, which strengthened the water H-bonding network around the Br atom. However, no evidence was found to support Hu *et al*'s[86] hypothesis that more "protective" water molecules should be present in the W2 position for U than for BrU.

Van Mourik *et al*. had employed a rather coarse-grained decomposition of their energies into simple base–water and water–water interaction terms, which were not further partitioned into physically meaningful discrete types of interaction (other than by qualitative visual inspection of the water networks). However, the origin of the aqueous stabilisation of minor tautomers of DNA bases can also be investigated on the level of orbital interactions. We have seen earlier that Hu *et al*. used the Natural Bond Orbital method. Likewise, Lukmanov *et al.* studied six tautomers of uracil and its 5- and 6-substituted derivatives (not including 5-BrU) in water, with the solvent simultaneously modelled by five explicit water molecules and PCM.[142] The diketo tautomer was the most stable, by

at least 24.5 kcal/mol, for all compounds, while the 4-enol-3-imide ("mutagenic") tautomer was the second most stable, except for 5-hydroxyuracil. NBO analysis showed that the diketo tautomers were stabilised by $n_N \rightarrow \pi^*$ and $n_N \rightarrow \sigma^*$ interactions between the lone pairs on NH groups and the antibonding orbitals of neighbouring bonds, and that this effect was sharply reduced in the enol tautomers (hence their lower stability), but partly compensated by increasing π-conjugation length. The role of the solvent in stabilising the enol tautomers of substituted uracils was twofold. "Specific" effects were those that required the inclusion of explicit water molecules, the most important being the formation of stronger H-bonds between O4-H4 and water when a 5-substituent was present, leading to greater aqueous stabilisation of the enol tautomers of 5-substituted uracils than those of uracil itself. In NBO terms, this consisted of $n_O \rightarrow \sigma^*_{OH}$ and $n_O \rightarrow \sigma^*_{NH}$ donor–acceptor interactions. "Non-specific" effects mostly involved the interaction of the dipoles of the tautomers with the electric field of the continuum solvent: the overall dipole moments were found to correlate well with the hydration energies. Hence both explicit *and* implicit solvation were deemed necessary to obtain correct quantitative information about relative tautomeric stabilities.

The cluster studies by van Mourik *et al.* had provided intriguing evidence that solvation in water considerably increases the population of the "mutagenic" tautomer of BrU, which possibly even becomes the dominant species. However, as mentioned above, quite a wide range of tautomerisation energies was found for both BrU and U, depending on cluster size and choice of functional. In general, the older, simpler functionals – BLYP (a member of the GGA class) and B3LYP (a hybrid GGA) – found a greater stabilisation of enolised BrU than the more recent, heavily parameterised functional M05-2X (a hybrid-meta GGA). Since BLYP and B3LYP are known to describe dispersion poorly or not at all,[119, 120, 143, 144] while M05-2X is specifically designed to include dispersion,[111] it seems likely that the latter should give more accurate results for this system. If so, the absolute value of the stabilisation of BrU-enol$(H_2O)_{50/100}$ over BrU-keto$(H_2O)_{50/100}$ may be quite small. In any case, it is desirable to have reliable quantitative data which converge to some limit.

We therefore decided to extend the studies by van Mourik *et al*. First of all, we opted to calculate single-point energies of the optimised clusters using the more recently developed functional M06-2X,[112] a re-parameterisation of M05-2X. We also calculated single-points with the local exchange functional M06-L,[145] developed in the same laboratory, as well as with PBE0.[146] This latter functional is a hybrid variant of PBE,[109] which itself is among the few DFT functionals that are able to capture some part of the dispersion energy without either parameterisation or post-processing corrections.[117] Subsequently, we carried out full geometry re-optimisations of the B3LYP-optimised complexes using M05-2X, M06-2X, M06-L and PBE0.

3.2. Definition of energetic terms

A cluster of 50 water molecules embedding either the keto or enol tautomer of a base is termed a *complex*, and is denoted BW50, where B = U or BrU and W = water. The tautomerisation energy of a solvated base is then defined as the difference between the formation energies of the enol-complex and the keto-complex of that base. The formation energies will now be defined.

In the following energetic terms (where E is electronic potential energy), the subscripts indicate the molecular system with which the energy is associated (either a single molecule, the cluster of 50 water molecules, or the entire base+water complex); the superscripts in curly brackets indicate the basis set in which the calculation is performed (either that of a single molecule or the entire complex); and the

round brackets indicate whether the molecular system has the geometry which it adopts in the optimised complex or its isolated (gas-phase) geometry.

The counterpoise-corrected formation energy of a complex, $\Delta E^{\mathrm{CP}}_{\mathrm{BW50}}$, is given by

$$\Delta E^{\mathrm{CP}}_{\mathrm{BW50}} = E^{\{\mathrm{BW50}\}}_{\mathrm{BW50}}(\mathrm{BW50}) - E^{\{\mathrm{BW50}\}}_{\mathrm{B}}(\mathrm{BW50}) - \sum_{i=1}^{50} E^{\{\mathrm{BW50}\}}_{\mathrm{Wi}}(\mathrm{BW50}) + E^{\mathrm{def}}_{\mathrm{B}} + E^{\mathrm{def}}_{\mathrm{W}} + \Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}} \quad (3.1)$$

where $E^{\{\mathrm{BW50}\}}_{\mathrm{BW50}}(\mathrm{BW50})$ is the total energy of the optimised complex calculated in the basis set of the entire complex, $E^{\{\mathrm{BW50}\}}_{\mathrm{B}}(\mathrm{BW50})$ is the energy of the base (in its complex-optimised geometry) calculated in the basis set of the entire complex, $E^{\{\mathrm{BW50}\}}_{\mathrm{Wi}}$ (BW50) are the energies of the water molecules (in their complex-optimised geometries) calculated in the basis set of the entire complex, $E^{\mathrm{def}}_{\mathrm{B}}$and $E^{\mathrm{def}}_{\mathrm{W}}$ are the deformation energies of the base and water, respectively, and $\Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}}$ is the gas-phase tautomerisation energy of the base.

It is the use of the basis set of the *entire* complex in calculating the isolated monomer energies that accomplishes the counterpoise correction.

The deformation energies, which are positive (unfavourable), result from the change in each monomer's geometry when passing from the gas phase to the optimised complex. They are calculated in the bare basis set of each monomer, and are given by

$$E^{\mathrm{def}}_{\mathrm{X}} = E^{\{\mathrm{X}\}}_{\mathrm{X}}(\mathrm{BW50}) - E^{\{\mathrm{X}\}}_{\mathrm{X}}(\mathrm{X}) \qquad \mathrm{X = B\ or\ W_i} \quad (3.2)$$

The gas-phase tautomerisation energy is the energy difference between a given tautomer of a base (optimised in the gas phase) and the keto tautomer of that base (also optimised in the gas phase):

$$\Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}} = E^{\{\mathrm{B}\}}_{\mathrm{B}}(\mathrm{B}) - E^{\{\mathrm{B}\}}_{\mathrm{B}}(\mathrm{B_{keto}}) \quad (3.3)$$

By definition, this energy is zero for the keto tautomer of each base. For the enol tautomers it is always positive. It is defined in this way because the keto tautomer is assumed to be the starting point for the tautomeric reaction.

The tautomerisation energy of each base is then given by

$$\Delta E_{\mathrm{keto\rightarrow enol}} = \Delta E^{\mathrm{CP}}_{\mathrm{B(enol)W50}} - \Delta E^{\mathrm{CP}}_{\mathrm{B(keto)W50}} \quad (3.4)$$

The tautomerisation energy can, in turn, be decomposed into five contributions, each of which is a difference between equivalent terms in the keto- and enol-complexes, combining to give the total keto-enol energy difference:

$$\Delta E_{\mathrm{keto\rightarrow enol}} = \Delta E^{\mathrm{def}}_{\mathrm{B}} + \Delta E^{\mathrm{def}}_{\mathrm{W}} + \Delta\Delta E_{\mathrm{B-W50}} + \Delta\Delta E_{\mathrm{W50}} + \Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}} \quad (3.5)$$

where

$$\Delta E^{\mathrm{def}}_{\mathrm{X}} = E^{\mathrm{def}}_{\mathrm{X(enol)}} - E^{\mathrm{def}}_{\mathrm{X(keto)}} \qquad \mathrm{X = B\ or\ W} \quad (3.6)$$

are the contributions of the base and water deformation energies,

$$\Delta\Delta E_{\mathrm{B-W50}} = E^{\{\mathrm{BW50}\}}_{\mathrm{B(enol)W50}}(\mathrm{B(enol)W50}) - E^{\{\mathrm{BW50}\}}_{\mathrm{B(enol)}}(\mathrm{B(enol)W50}) - E^{\{\mathrm{BW50}\}}_{\mathrm{W50(enol)}}(\mathrm{B(enol)W50}) - E^{\{\mathrm{BW50}\}}_{\mathrm{B(keto)W50}}(\mathrm{B(keto)W50}) + E^{\{\mathrm{BW50}\}}_{\mathrm{B(keto)}}(\mathrm{B(keto)W50}) + E^{\{\mathrm{BW50}\}}_{\mathrm{W50(keto)}}(\mathrm{B(keto)W50}) \quad (3.7)$$

is the contribution of the base–water interaction energies,

$$\Delta\Delta E_{\mathrm{W50}} = E^{\{\mathrm{BW50}\}}_{\mathrm{W50(enol)}}(\mathrm{B(enol)W50}) - \sum_{i=1}^{50} E^{\{\mathrm{BW50}\}}_{\mathrm{Wi(enol)}}(\mathrm{B(enol)W50}) - E^{\{\mathrm{BW50}\}}_{\mathrm{W50(keto)}}(\mathrm{B(keto)W50}) + \sum_{i=1}^{50} E^{\{\mathrm{BW50}\}}_{\mathrm{Wi(keto)}}(\mathrm{B(keto)W50}) \quad (3.8)$$

is the contribution of the water–water interaction energies, and $\Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}}$ was defined earlier.

3.3. Single-point calculations of 50-water clusters

Our first approach was to calculate M06-2X/6-31G(d,p), M06-L/6-31G(d,p) and PBE0/6-31G(d,p) single-point formation energies of the 50-water keto and enol complexes of BrU and U, in the geometries already optimised by van Mourik *et al.* This would lead directly to tautomerisation energies. We chose to use the B3LYP/6-31G(d,p) geometries.

All calculations were performed using Gaussian 09. To decompose the formation energies into the contributions shown in Equation 3.1, the counterpoise procedure was split into separate parts. For each complex, a single calculation provided the total complex energy in the complete basis set ($E^{\{\mathrm{BW50}\}}_{\mathrm{BW50}}(\mathrm{BW50})$). Separate calculations were run for the individual energies of the base and each water molecule in the complete basis set ($E^{\{\mathrm{BW50}\}}_{\mathrm{B}}(\mathrm{BW50})$) and $\sum_{i=1}^{50} E^{\{\mathrm{BW50}\}}_{\mathrm{Wi}}(\mathrm{BW50})$), the individual energies of the base and each water molecule in their own basis sets ($E^{\{\mathrm{B}\}}_{\mathrm{B}}(\mathrm{BW50})$) and $\sum_{i=1}^{50} E^{\{\mathrm{W}\}}_{\mathrm{Wi}}(\mathrm{BW50})$), and the total energy of the 50 waters in the complete basis set ($E^{\{\mathrm{BW50}\}}_{\mathrm{W50}}(\mathrm{BW50})$). Combined with optimisations of both tautomers of each base in the gas phase, and of water in the gas phase, this allowed derivation of the gas-phase tautomerisation energies (shown in Table 3.1) and the deformation energies, as well as the base–water and water–water interaction terms. The energies of the base, the individual waters, and the 50 waters in the complete basis set were calculated by specifying that the remaining atoms were "ghost atoms", which supply basis functions but no electrons or nuclei. By splitting up the counterpoise procedure in this way, the "Counterpoise" keyword was not needed.

3.3.1. M06-2X

For M06-2X, the counterpoise-corrected formation energies and tautomerisation energies are shown in Table 3.3. Table 3.3 also shows the magnitude of the counterpoise correction (BSSE) for each formation energy, and the contributions of base–water and water–water interactions to each formation energy. The decomposition of the tautomerisation energies, as shown in Equation 3.5, is given in Table 3.4.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−612.6	−2.5	−186.9	−66.9	−592.9
BrU(enol)W50	−615.0		−174.9	−78.4	−600.2
U(keto)W50	−630.1	8.7	−173.4	−61.9	−614.0
U(enol)W50	−621.4		−168.9	−83.2	−603.4

Table 3.3. Formation energies, tautomerisation energies, basis set superposition errors, base−water interaction energies and water−water interaction energies, in kcal/mol, of BW50 at the M06-2X/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	1.2	3.8	−11.5	−7.3	11.4
UW50	5.7	2.9	−21.3	10.6	10.8

Table 3.4. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base−water interaction, water−water interaction and gas-phase tautomerisation, at the M06-2X/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

Table 3.3 shows that the tautomerisation of BrU is favoured by −2.5 kcal/mol, while for U it is disfavoured by +8.7 kcal/mol. The corresponding M05-2X/6-31G(d,p) energies from Ref.[88] (using the same geometries) were −3.3 kcal/mol and +7.9 kcal/mol. Hence the use of M06-2X rather than M05-2X makes only a small difference. The individual energy contributions from M05-2X (available in Ref.[88]) are also all relatively unaffected. The crucial factor is still the water−water interaction, which favours tautomerisation of BrU but not U. The base−water interaction actually favours tautomerisation for U more than for BrU, like with M05-2X.

3.3.2. M06-L

The results from M06-L/6-31G(d,p) are given in Tables 3.5 and 3.6.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−559.0	4.2	−157.4	−67.5	−546.9
BrU(enol)W50	−554.8		−147.0	−76.1	−552.0
U(keto)W50	−574.7	14.5	−144.9	−61.4	−567.0
U(enol)W50	−560.2		−141.3	−80.0	−555.4

Table 3.5. Formation energies, tautomerisation energies, basis set superposition errors, base−water interaction energies and water−water interaction energies, in kcal/mol, of BW50 at the M06-L/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	1.2	3.8	−8.6	−5.1	13.0
UW50	5.9	2.8	−18.6	11.6	12.7

Table 3.6. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base−water interaction, water−water interaction and gas-phase tautomerisation, at the M06-L/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

In contrast to M06-2X, M06-L returns a positive (unfavourable) tautomerisation energy for both bases. However, the penalty is much higher for U than for BrU – in fact, for U, tautomerisation is less favourable in the 50-water cluster (14.5 kcal/mol) than in the gas phase (12.7 kcal/mol). For BrU, when the M06-L energies are compared to those from M05-2X and M06-2X, it can be seen that three

terms are responsible for the shift in tautomerisation energy from favourable to unfavourable: the base–water and water–water contributions become less strongly favourable, and the gas-phase tautomerisation becomes more strongly unfavourable.

3.3.3. PBE0

The results from PBE0/6-31G(d,p) are given in Tables 3.7 and 3.8.

	$\Delta E^{\mathrm{CP}}_{\mathrm{BW50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−591.3	−12.3	−202.6	−50.0	−591.9
BrU(enol)W50	−603.6		−190.2	−66.4	−605.2
U(keto)W50	−616.3	4.2	−189.0	−48.0	−617.1
U(enol)W50	−612.1		−184.1	−74.8	−607.0

Table 3.7. Formation energies, tautomerisation energies, basis set superposition errors, base–water interaction energies and water–water interaction energies, in kcal/mol, of BW50 at the PBE0/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

	$\Delta E^{\mathrm{def}}_{\mathrm{B}}$	$\Delta E^{\mathrm{def}}_{\mathrm{W}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E^{\mathrm{gas}}_{\mathrm{keto\rightarrow enol}}$
BrUW50	1.2	3.7	−16.4	−13.3	12.5
UW50	5.9	2.9	−26.8	10.1	12.0

Table 3.8. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base–water interaction, water–water interaction and gas-phase tautomerisation, at the PBE0/6-31G(d,p)//B3LYP/6-31G(d,p) level of theory.

PBE0/6-31G(d,p) favours tautomerisation of BrU by −12.3 kcal/mol, but disfavours tautomerisation of U by +4.2 kcal/mol. These values are close to the B3LYP/6-31G(d,p) energies obtained by van Mourik *et al.*, i.e. −13.3 and +3.8 kcal/mol, respectively. The relative contributions to the tautomerisation energies are also very similar between the two functionals – for example, with PBE0/6-31G(d,p), the water–water term $\Delta\Delta E_{\mathrm{W50}}$ = −13.3 kcal/mol for BrU and +10.1 kcal/mol for U, compared to −13.0 kcal/mol and +10.2 kcal/mol, respectively, with B3LYP/6-31G(d,p).

3.4. Geometry optimisations of 50-water clusters

The formation energy of a complex is likely to be highly sensitive to its geometry. We have already seen that the water–water interaction energies, which are determined by the structure of the water network, have a decisive influence on the relative energies of the tautomers of a base. For such large systems, however, a geometry that represents a local energy minimum for one functional may be quite far from the nearest minimum on the potential energy surface calculated by a different functional. Therefore, to obtain more accurate formation energies from the Minnesota functionals and PBE0, we have used them to fully re-optimise the BW50 clusters, again starting from the B3LYP/6-31G(d,p) geometries. As before, these calculations were performed with Gaussian 09.

We initially used the default optimisation algorithm in that program, the Berny algorithm,[133] but the optimisations usually exceeded the allowed number of steps, having been trapped in a loop alternating between two geometries. We therefore switched to the DL-Find algorithm,[147] implemented via the program ChemShell.[148] This program assumes the task of determining the next move in the geometry optimisation, using DL-Find, while delegating the actual energy calculations to Gaussian 09. Our

optimisations exploited the hybrid delocalised internal coordinates (HDLC) module within ChemShell's implementation of DL-Find. This method partitions the system into fragments, which are internally defined using primitive internal coordinates (as in the Berny algorithm), but are then coupled together using the Cartesian coordinates of each fragment. This simplifies the transformation from Cartesian coordinates (the user input) to internal coordinates (necessary for determining the next optimisation step), and thus reduces the scaling of the computational cost of geometry optimisation with respect to system size. We chose the individual base and water molecules as the fragments.

The geometry optimisations employed the M05-2X, M06-2X, M06-L and PBE0 functionals, combined with the 6-31G(d,p) basis set. For all four complexes, starting from the B3LYP/6-31G(d,p) geometries obtained by van Mourik *et al*., a stationary point on the potential energy surface using each new functional was successfully located. The formation energies were then calculated in the same way as before. The calculation and decomposition of the tautomerisation energies also proceeded in the same way. The results are shown in Tables 3.9–3.16. A discussion of the energies and geometries then follows.

3.4.1. M05-2X and M06-2X

The optimised formation and tautomerisation energies from M05-2X are given in Tables 3.9 and 3.10, respectively, while those from M06-2X are given in Tables 3.11 and 3.12, respectively. The two functionals will be discussed together.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B\text{–}W50}}$	ΔE_{W50}
BrU(keto)W50	−607.7 (−600.2)	7.0 (−3.3)	−215.3 (−189.4)	−50.2 (−64.1)	−593.2 (−588)
BrU(enol)W50	−600.7 (−603.5)		−201.8 (−177.7)	−71.2 (−76.8)	−589.4 (−595)
U(keto)W50	−617.0 (−619.0)	7.6 (7.9)	−195.3 (−176.1)	−53.3 (−59.9)	−601.2 (−609)
U(enol)W50	−609.5 (−611.1)		−182.2 (−171.7)	−90.7 (−82.4)	−585.0 (−599)

Table 3.9. Formation energies, tautomerisation energies, basis set superposition errors, base–water interaction energies and water–water interaction energies, in kcal/mol, of BW50 optimised at the M05-2X/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries, from Ref. [88]

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B\text{–}W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	10.0 (1.2)	2.9 (4.1)	−21.0 (−12.7)	3.8 (−7.4)	11.3 (11.4)[iv]
UW50	12.2 (6.0)	5.9 (3.1)	−37.4 (−22.5)	16.2 (10.7)	10.7 (10.7)[iv]

Table 3.10. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base–water interaction, water–water interaction and gas-phase tautomerisation, optimised at the M05-2X/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries, from Ref. [88]

[iv] There is a slight discrepancy between the M05-2X/6-31G(d,p) gas-phase tautomerisation energies reported in Ref. [88] and those used in this thesis.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−616.9 (−612.6)	2.0 (−2.5)	−225.2 (−186.9)	−59.2 (−66.9)	−595.1 (−592.9)
BrU(enol)W50	−614.9 (−615.0)		−228.8 (−174.9)	−70.9 (−78.4)	−599.6 (−600.2)
U(keto)W50	−621.7 (−630.1)	10.7 (8.7)	−217.4 (−173.4)	−58.2 (−61.9)	−600.2 (−614.0)
U(enol)W50	−611.0 (−621.4)		−201.5 (−168.9)	−156.7 (−83.2)	−579.0 (−603.4)

Table 3.11. Formation energies, tautomerisation energies, basis set superposition errors, base−water interaction energies and water−water interaction energies, in kcal/mol, of BW50 optimised at the M06-2X/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries.

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	5.9 (1.2)	0.8 (3.8)	−11.6 (−11.5)	−4.5 (−7.3)	11.4 (11.4)
UW50	64.7 (5.7)	12.6 (2.9)	−98.5 (−21.3)	21.1 (10.6)	10.8 (10.8)

Table 3.12. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base−water interaction, water−water interaction and gas-phase tautomerisation, optimised at the M06-2X/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries.

Tables 3.9 and 3.11 show that, when the clusters were optimised with either M05-2X or M06-2X, the *keto* tautomer of BrU was preferred (like in the gas phase), as was the keto tautomer of U. This reverses the central finding of van Mourik *et al*. However, the tautomerisation energies were quantitatively different between these two functionals. With M05-2X, the two energies were very close: 7.6 kcal/mol for U and 7.0 kcal/mol for BrU. With M06-2X, the tautomerisation energy of U (10.7 kcal/mol) was substantially less favourable than that of BrU (2.0 kcal/mol). With both functionals, the base−water interaction favoured the enol form of each base, but by a larger amount for U. With both functionals, the water−water interaction was more conducive to tautomerisation of BrU than of U; however, with M05-2X it actually disfavoured tautomerisation of each base overall, while with M06-2X it favoured the enol form of BrU and the keto form of U.

All the optimisations with both functionals resulted in noticeable rearrangements of the water networks. The water−water interactions had been crucial to the preference for the enol tautomer of BrU reported by van Mourik *et al*. It should therefore be expected that structural changes to the water network would have significant energetic consequences for the tautomeric shifts, and this proved so. Table 3.9 shows that for BrU(enol)W50, U(keto)W50 and U(enol)W50, the water−water interaction is actually *weaker* following re-optimisation with M05-2X and M06-2X, while for BrU(keto)W50 it becomes stronger. This is why the water−water interaction is less conducive to the tautomerisation of BrU at these new geometries. However, the amount by which the base−water interaction favours tautomerisation of U over that of BrU also increases following optimisation with M05-2X and M06-2X. Furthermore, the base−water interactions are again weaker in the M05-2X and M06-2X geometries of BrU(keto)W50, BrU(enol)W50 and U(keto)W50.

Therefore for two complexes – BrU(enol)W50 and U(keto)W50 – *both* components of the M05-2X and M06-2X interaction energies (base–water and water–water) became weaker following optimisation with those functionals!

Overall, Tables 3.9 and 3.11 show that for three complexes – BrU(enol)W50, U(keto)W50 and U(enol)W50 – the final, counterpoise-corrected formation energies are *less* favourable in the optimised M05-2X and M06-2X geometries than those obtained with the same functionals at the original geometries.

These seemingly paradoxical results are due to basis set superposition error. For all four complexes, the quantity $E_{\mathrm{BW50}}^{\{\mathrm{BW50}\}}(\mathrm{BW50})$, the total *uncorrected* energy of the entire complex in the complete (superposed) basis set, is more negative in the optimised geometries than in the starting geometries. However, once the spurious negative contribution of the BSSE is discarded, the total *counterpoise-corrected* energies of the BrU(enol)W50, U(keto)W50 and U(enol)W50 complexes are actually less negative in the final geometries than in the starting geometries. This is because basis set superposition has biased the geometry optimisations towards a closer mutual approach of the monomers, resulting in geometries that increase the BSSE but weaken the genuine attractive interactions between the monomers.

The substantial BSSE is clearly due to the use of a relatively small basis set, 6-31G(d,p), a choice necessitated by the size of the systems under study. It is also possible that M05-2X and M06-2X are more susceptible to BSSE than B3LYP is. Mardirossian and Head-Gordon showed that some of the Minnesota functionals converge very slowly with respect to basis set size, and thus experience substantial BSSE even with quadruple- and quintuple-ζ basis sets.[149] This behaviour was shown to be due to the inhomogeneity correction factor (a term that corrects for the local spin density approximation) – the large number of parameters in this term makes it oscillate wildly with respect to the local kinetic energy ratio. However, with double-ζ basis sets, the BSSE appeared to be identical for B3LYP, M05-2X, M06-2X, M06-L and PBE according to that study.

The large BSSE casts some doubt on the validity of these optimised geometries; presumably, they would have been markedly different had the counterpoise correction been applied *during* the optimisations. That, however, would have incurred a great computational expense.

Note that during the optimisation of U(enol)W50 with M06-2X (but not M05-2X), the O4-hydroxyl proton transferred to a nearby water molecule, generating an H_3O^+ ion and anionic U^-. This resulted in a highly favourable base–water interaction (−157 kcal/mol, compared to typically −70 kcal/mol), balanced by a highly unfavourable base deformation term (+68 kcal/mol, compared to typically +7 kcal/mol). This effect – possibly driven by BSSE – explains why the relative contributions of base deformation and base–water interaction to the tautomerisation energy of UW50 are very different between M05-2X and M06-2X (cf. Tables 3.10 and 3.12). With BrUW50, on the other hand, optimisation using M06-2X did not cause any such effect; in energetic terms, the base–water interaction continued to favour tautomerisation by almost exactly the same amount as before, while the water–water interaction contributed less stability to the enol tautomer than before, resulting in the positive overall tautomerisation energy.

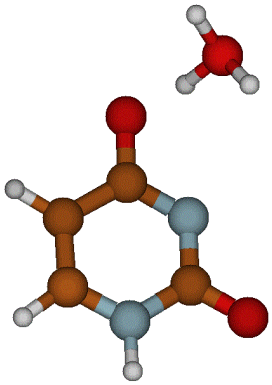


Figure 3.2: Complex between anionic U^- and a water molecule in U-enol$(H_2O)_{50}$ optimised with M06-2X/6-31G(d,p). The other waters are omitted for clarity.

Some more structural observations can be made. Van Mourik *et al.* noted that the 5-position in their UW50 clusters was on the hydrophobic side of the base, i.e. there was no interaction between the H5 atom and any water molecules.[88] In their B3LYP-optimised structures, the shortest H5---O_w distance was 2.6 Å in U(keto)W50 and 2.3 Å in U(enol)W50. Our re-optimisation with M05-2X and M06-2X does not qualitatively change this: with M05-2X, the equivalent optimised distances are 2.6 Å and 2.4 Å, while with M06-2X they are 2.7 Å and 2.4 Å.

In their BrU(keto)W50 cluster, van Mourik *et al.* observed a short Br---H_w contact (2.8 Å), but did not deem it energetically significant. Our optimisation with M05-2X increases this distance to 3.0 Å, and also decreases the linearity of the Br---H_wO_w angle from 143° to 123°; on the other hand, it results in another Br---H_w contact with a distance of 2.8 Å and an angle of 122°. Optimisation with M06-2X increases the distance of the Br---H_wO_w contact to 2.9 Å and reduces the angle to 118°, while resulting in another contact, with distance 2.7 Å and angle 117°. Likewise, the BrU(enol)W50 cluster from B3LYP contains a Br---H_w contact of distance 2.9 Å with an angle of 122°. M05-2X merely increases the angle to 131°, while M06-2X also increases the distance to 3.1 Å and decreases the angle to 118°; other short contacts exist as well, but have very acute Br---H_wO_w angles. Therefore Br---H_w contacts do not seem to contribute much to the stability of BrU(W50) complexes.

Bromine and water can also take part in *halogen bonds*: attractive, non-covalent interactions between a carbon-bonded halogen (C-X) and an electronegative oxygen (O-Y, where Y = H in the case of water).[150, 151] These interactions are due to a small electropositive region on the outer tip of the halogen atom, known as the σ-hole, which is attracted to the electronegative oxygen. Halogen bonds are characterised by linear C-X---O bond angles (~165°) and short X---O distances (≤ 3.37 Å when X = Br). Being essentially electrostatic in nature, halogen bonds can be adequately described by DFT, provided at least a double-ζ basis set is used.[152] Rastogi *et al.* calculated at the B3LYP/6-311++G(2d,p) level that the optimum geometry of monohydrated BrU contained a linear C-Br---O_w bond angle, with a distance of 3.1 Å, i.e. a possible halogen bond.[153]

The starting structure of BrU(keto)W50 contains a candidate halogen bond, with a C-Br---O_w angle of 168° and a Br---O_w distance of 3.1 Å. Optimisation with M05-2X increases this angle to 175° and reduces the distance to 2.9 Å, while the corresponding figures for M06-2X are 166° and 2.9 Å. The starting structure of BrU(enol)W50 also contains a candidate halogen bond, with an angle of 157° and distance of 3.2 Å. These values become 172° and 2.9 Å following optimisation with M05-2X. Optimisation with M06-2X appears to create two somewhat bifurcated halogen bonds, the first with angle 151° and distance 3.0 Å, the second with angle 154° and distance 3.0 Å. Although the

decreases in bond length and increases in linearity might suggest favourable interactions, it is difficult to argue that these contacts are actually stabilising, since the base–water interaction energies become less favourable for both tautomers of BrU following optimisation with either method.

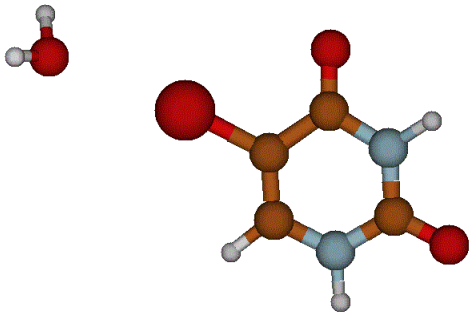


Figure 3.3: Possible halogen bond between Br and O_w in BrU-keto$(H_2O)_{50}$ optimised with M05-2X/6-31G(d,p). The other waters are omitted for clarity.

3.4.2. M06-L

The results from M06-L are shown in Tables 3.13 and 3.14.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−570.7 (−559.0)	13.5 (4.2)	−179.2 (−157.4)	−56.2 (−67.5)	−547.2 (−546.9)
BrU(enol)W50	−557.1 (−554.8)		−178.4 (−147.0)	−68.0 (−76.1)	−538.4 (−552.0)
U(keto)W50	−586.0 (−574.7)	24.8 (14.5)	−163.3 (−144.9)	−64.5 (−61.4)	−556.4 (−567.0)
U(enol)W50	−561.2 (−560.2)		−153.7 (−141.3)	−73.7 (−80.0)	−535.3 (−555.4)

Table 3.13. Formation energies, tautomerisation energies, basis set superposition errors, base–water interaction energies and water–water interaction energies, in kcal/mol, of BW50 optimised at the M06-L/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries.

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	4.7 (1.2)	−1.2 (3.8)	−11.8 (−11.5)	8.8 (−7.3)	13.0 (11.4)
UW50	1.4 (5.7)	−1.2 (2.9)	−9.2 (−21.3)	21.1 (10.6)	12.7 (10.8)

Table 3.14. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base–water interaction, water–water interaction and gas-phase tautomerisation, optimised at the M06-L/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but the B3LYP/6-31G(d,p) geometries.

Optimisation with M06-L has a dramatic result: tautomerisation of both BrU and U becomes more unfavourable than it was with M06-L single-points in the B3LYP geometries, and in fact even more unfavourable than in the gas phase. For BrU, this is mostly caused by changes in the water–water interaction, which favour the enol tautomer in the B3LYP geometries but the keto tautomer in the M06-L geometries. For U, the very highly positive tautomerisation energy (24.8 kcal/mol) is due to both the water–water contribution becoming positive for tautomerisation, and the base–water contribution becoming less negative for tautomerisation.

Unlike with M05-2X and M06-2X, in M06-L the BSSE is not solely responsible for the decrease in energy during optimisation. Even after the counterpoise correction is applied, the total energies of all four complexes are more negative in the optimised geometries than in the starting geometries. It can be seen from Tables 3.9, 3.11 and 3.13 that the total BSSE is smaller by around 50 kcal/mol for M06-L (in the optimised geometries) than for the other two methods. This is presumably due to its local character, which discourages the overlap of basis functions. However, the BSSE is still significant in M06-L – note that for both the enol-complexes, the optimised energies are less than 5 kcal/mol below the starting energies after counterpoise is applied, while for the keto-complexes the equivalent difference is around 12 kcal/mol.

For U(keto)W50, the shortest H5---O_w contact remains 2.6 Å after optimisation. For U(enol), the shortest H5---O_w contact increases from 2.3 to 2.4 Å.

For BrU(keto)W50, the shortest Br---H_w contact remains 2.8 Å after optimisation, with a Br---H_wO_w angle of 128°. The potential halogen bond in the starting structure experiences an increase in the C-Br---O_w angle from 168° to 173° and a decrease in the Br---O_w distance from 3.1 Å to 2.9 Å. For BrU(enol)W50, the optimised structure contains numerous short Br---H_w contacts, the shortest of which has a length of 2.8 Å and a Br---H_wO_w angle of 110°. The angle of the potential halogen bond, however, changes from 157° to 145°, and its distance changes from 3.2 Å to 3.1 Å. This reduction of the angle away from linearity indicates that it is not a halogen bond. Both complexes have less favourable base–water interaction terms in the optimised geometries than in the starting geometries, as with M05-2X and M06-2X.

3.4.3. PBE0

The results from PBE0 are shown in Tables 3.15 and 3.16.

	$\Delta E_{\mathrm{BW50}}^{\mathrm{CP}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}$	BSSE	$\Delta E_{\mathrm{B-W50}}$	ΔE_{W50}
BrU(keto)W50	−593.2 (−591.3)	−6.7 (−12.3)	−216.2 (−202.6)	−59.2 (−50.0)	−595.1 (−591.9)
BrU(enol)W50	−599.9 (−603.6)		−196.3 (−190.2)	−66.5 (−66.4)	−599.7 (−605.2)
U(keto)W50	−613.3 (−616.3)	3.7 (4.2)	−194.1 (−189.0)	−47.9 (−48.0)	−611.3 (−617.1)
U(enol)W50	−609.6 (−612.1)		−189.0 (−184.1)	−75.8 (−74.8)	−601.5 (−607.0)

Table 3.15. Formation energies, tautomerisation energies, basis set superposition errors, base–water interaction energies and water–water interaction energies, in kcal/mol, of BW50 optimised at the PBE0/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries.

	$\Delta E_{\mathrm{B}}^{\mathrm{def}}$	$\Delta E_{\mathrm{W}}^{\mathrm{def}}$	$\Delta\Delta E_{\mathrm{B-W50}}$	$\Delta\Delta E_{\mathrm{W50}}$	$\Delta E_{\mathrm{keto\rightarrow enol}}^{\mathrm{gas}}$
BrUW50	3.8 (1.2)	4.8 (3.7)	−23.1 (−16.4)	−4.7 (−13.3)	12.5 (12.5)
UW50	6.6 (5.9)	3.2 (2.9)	−28.0 (−26.8)	9.8 (10.1)	12.0 (12.0)

Table 3.16. Decomposition of the tautomerisation energies of BW50 into the contributions, in kcal/mol, of base deformation, water deformation, base–water interaction, water–water interaction and gas-phase tautomerisation, optimised at the PBE0/6-31G(d,p) level of theory. In brackets are the corresponding energies using the same method but at the B3LYP/6-31G(d,p) geometries.

PBE0 returns the same qualitative overall result as the methods originally employed by van Mourik *et al.*:[88] tautomerisation is favourable for BrU but not U. The contributions are also qualitatively the same as with van Mourik *et al*'s BW50 clusters: base–water interaction favours tautomerisation of both bases, but that of U by more; water–water interaction favours tautomerisation of BrU but hinders that of U; base deformation hinders tautomerisation of both bases, but that of U by more; and water deformation hinders tautomerisation of both bases, but that of BrU by more.

However, the final geometries are again decisively influenced by BSSE. After the counterpoise correction is applied, all the complexes except BrU(keto)W50 have *less* favourable formation energies in the final geometries than in the initial geometries (see the values in brackets in the first column of Table 3.15). This is due to unfavourable changes in the water network: for all the complexes except BrU(keto)W50, the water–water interactions are weaker in the final state than in the starting state, while the base–water interactions are essentially unchanged. Only for BrU(keto)W50 do both interactions become stronger. In fact, the PBE0 geometry of BrU(keto)W50 is the only complex containing BrU in which the CP-corrected base–water interaction is significantly stronger (by 9.3 kcal/mol) after optimisation.[v]

This base–water strengthening in BrU(keto)W50 is possibly connected to the candidate halogen bond, which has an optimised bond angle of 174° (compared to 168° in the starting structure) and a bond length of 2.9 Å (cf. 3.1 Å in the starting structure). The shortest Br---H_w contact has a distance of 2.7 Å (decreased from 2.8 Å) and a Br---H_wO_w angle of 122° (decreased from 143°). These changes are consistent with a *slightly* stronger interaction between water and the bromine atom, but remember that the total change in the base–water interaction energy is quite large (9.3 kcal/mol). The corresponding optimised values in BrU(enol)W50 are 158° and 3.1 Å for the halogen bond (changed from 157° and 3.2 Å), and 2.9 Å and 123° for the Br---H_w contact (compared to 2.9 Å and 122°). These negligible changes are consistent with the essentially zero change in $\Delta E_{\mathrm{B-W50}}$ for that complex.

3.5. Discussion

Of all 16 cluster optimisations described above, only four resulted in significantly stronger (by ≥3.5 kcal/mol) base–water or water–water interactions, compared to the starting structures, after accounting for BSSE. One of these four was the base–water interaction in U(enol)W50 with M06-2X, which was clearly due to the ionic interaction between H_3O^+ and U^- in the final structure. The other three were base–water in U(enol)W50 with M05-2X, base–water in BrU(keto)W50 with PBE0, and water–water in BrU(keto)W50 with M05-2X. The observation that both interaction terms became weaker (or were essentially unchanged) in all four optimisations of BrU(enol)W50 can be related to

[v] The base–water interaction in BrU(enol)W50 is strengthened by a negligible 0.1 kcal/mol at the same level of theory (Table 3.15).

the fact that all four methods predicted enolisation of BrU to be less favourable in the optimised structures. Unfortunately, without a formal method of partitioning the interaction energy into physically meaningful terms (e.g. DFT-SAPT), it is difficult to establish the physical origin of these energy changes using the supermolecular method alone. In any case, the large influence of BSSE on the final geometries casts doubt on their validity.

However, the rare tautomer hypothesis of BrU mutagenicity does not require that the mutagenic tautomer be *absolutely* favoured over the canonical tautomer. Van Mourik *et al.*'s calculated reversal of the tautomeric preference of BrU in nanodroplets was a remarkable finding, but all that is actually required by the rare tautomer model is that the equilibrium population of the enol form, under physiological conditions, be appreciably higher for BrU than for U (or T). The keto form could still dominate overall.

Loeb and Kunkel estimated that, in the absence of repair apparatus, an error rate of one in 10^n base pairs required a free energy difference of $1.4n$ kcal/mol between correct and incorrect base pairs.[19] Equivalently, Lasken and Goodman[154] derived an energy difference of 2.7 kcal/mol for an error rate of one in 10^2.

Let us assume that the enolisation of aqueous BrU, though unfavourable, incurs a smaller energetic penalty than the enolisation of aqueous U. Then, BrU should form rare-tautomeric mispairs more often than U. Now, Loeb and Kunkel's equation implies that, for BrU to increase the error rate by 10^n (neglecting repair apparatus), the enolisation energy of BrU must be lower than that of U by a margin of $1.4n$ kcal/mol, assuming that all mispairs formed by either base are rare-tautomeric. Experimentally, the increase in error rate induced by BrU varies widely, as a function of many variables, but is generally in the range $10^2 - 10^4$.[53, 57, 61, 155] This, then, requires the enolisation energy of BrU to be 3 – 6 kcal/mol lower than that of U. However, since repair enzymes heavily reduce the mutation rate *in vivo*, this is probably an underestimate, representing the lower bound of the rare tautomer stabilisation needed to account for BrU-induced mutagenesis.

Note that Loeb and Kunkel's equation is not exact here (even if free rather than potential energies are available), because our clusters only contain single U or BrU bases, not base pairs. However, let us assume that BrU and U form hydrogen bonds with roughly the same strength, and likewise for BrU* and U*. Then, consider the case of AT→GC transitions, which are caused by substrate G (rather than A) mispairing with either U or BrU in the template. We have assumed that the energy difference between BrU*-G and BrU-A, and the energy difference between U*-G and U-A, differ from each other mostly due to the different enolisation energies of BrU and U, since the H-bonding patterns are the same. Hence the difference in base pair energies is approximately equal to the difference in enolisation energies. The same applies to the mispairing of substrate U and BrU, in competition with C, opposite template G: the difference between BrU*-G and C-G energies, and the difference between U*-G and C-G energies, should be roughly equal to the difference in enolisation energies of BrU and U.

Every method employed in this chapter, except for re-optimisation with M05-2X, found the tautomerisation of BrU to be more favourable than that of U by 8.7 kcal/mol or more. On this basis, we would expect BrU to induce mutations at at least 10^6 times the rate of U (8.7 / 1.4 = 6.2, cf. Loeb and Kunkel),[19] in the absence of repair enzymes. Allowing for the subsequent effect of those enzymes, this is arguably consistent with the experimental rates of $10^2 - 10^4$. Although it is disappointing that we did not have the time and resources to explore this line of enquiry further (e.g.

using different starting structures), and that the results we obtained were contaminated by BSSE, our findings in this chapter confirm that water approaching the bulk limit has a large and favourable effect on the tautomerisation of BrU compared to U, which cannot be predicted by implicit solvent models alone.

A natural follow-up question is to ask what size of water cluster is needed to appreciably change the gas-phase tautomeric preference of BrU. In particular, it would be productive to investigate smaller clusters (e.g. up to 10 water molecules), with the water in positions corresponding to the most commonly occupied hydration sites in DNA as established by crystallography.[45]

Chapter 4: CPMD study of tautomerism in a periodic water box

4.1. Background

It was shown in Chapter 3 that aqueous solvation reduces (or might even reverse) the energy difference between the keto and enol tautomers of 5-bromouracil, thereby increasing the equilibrium population of the enol tautomer. This conclusion was based on geometry optimisations of BrU in static water clusters. However, since liquid structure is dynamic, it is desirable to have dynamical information on the energetics of BrU and U tautomerisation, to further assess the likelihood that BrU causes mutagenesis through its enol tautomer.

Various dynamical studies of water-assisted DNA base tautomerisation can be found in the literature. In 2008, Fogarasi ran several hundred simulation trajectories of the "canonical" (amino-oxo) tautomer of cytosine complexed with a single water molecule.[156] These simulations employed a hybrid classical/quantum setup, in which the atoms moved classically but the potential energy surface was calculated on-the-fly with B3LYP. In one single trajectory, the water molecule acted as a bridge in the synchronous, concerted transfer of a hydrogen atom from the amino group to N3. This intramolecular proton movement generated the "rare" imino-oxo tautomer of cytosine, which is not observed in the gas phase but is believed to be stabilised by water. This study therefore provided a qualitative mechanistic account of how water may take part in tautomerisation, but did not provide quantitative data.

In 2011, Furmanchuk *et al.* performed simulations of the natural DNA bases solvated by 54–74 water molecules in a quantum (CPMD) setup, and separate classical simulations with the bases solvated by around 1500 waters.[137] The goal was to measure the abundance and lifetimes of bifurcated water bridges: chains of alternating H and O atoms leading from an N-H or O-H moiety to a non-hydrogenated N or O on the same base, formed by base–water hydrogen bonds (as well as water–water H-bonds in the case of bridges containing two waters). These bridges were deemed necessary for water-assisted tautomerisation according to models that they had derived from static *ab initio* calculations. They found that both unimolecular and bimolecular water bridges were rare, in the classical as well as the quantum simulations – usually existing for much less than 50% of the simulation time. In thymine, the most common bridge was to O4, while in cytosine, it was to N3 (both leading to the potentially mutagenic tautomers of these bases). In adenine, however, the most common bridge was to N7 (not forming a mutagenic tautomer). Using the existence times of these bridges to apply a quantitative correction to *ab initio* rate constants of tautomerisation, they found that the lack of bridges considerably reduced the population of rare tautomers that could be formed on a relevant timescale, especially for guanine. They concluded that water-assisted proton transfer could not be a significant source of mutagenic tautomers, except for cytosine. However, they did not directly calculate rate constants from their CPMD data.

Kinetics as well as thermodynamics may be relevant here. Gorb *et al.* have shown that a crucial effect of the interaction of water with DNA bases is to reduce the height of the free energy barriers separating different tautomers.[157] In their example (based on MP2 calculations), a water molecule

located between N1-H1 and O6 of guanine increases the energy of the mutagenic tautomer relative to the canonical tautomer, but also lowers the barrier height between them. By acting as a conduit for proton transfer, water stabilises the transition state between the tautomers. This accelerates tautomerisation – i.e. it increases the rate constant of the forward reaction in the major↔minor tautomeric equilibrium – thus increasing the likelihood that a significant fraction of mutagenic tautomers can be formed on the relatively short timescale of DNA replication, notwithstanding their lower equilibrium population. A similar conclusion was drawn by Markova *et al.* in an MP2 study of the oxo-hydroxy tautomerism of 5-fluorouracil, a cytostatic analogue of thymine and 5-bromouracil.[158]

Let us assume that the enolisation of BrU involves the water-assisted intramolecular transfer of a single proton from N3 to O4 (as opposed to Löwdin's[26] double proton tunnelling model). A suitable source of dynamical information about this process – and the corresponding process for uracil – is CPMD.

In 2011, van Mourik and Gaigeot performed a CPMD study of BrU and U in a periodic water box.[89] Their aim was to compare the free energies of deprotonation at N3 for these two molecules. The free energies were to be calculated according to the method of Sprik and Ciccotti:[159] a reaction coordinate is specified, and intermediate points along this coordinate are defined by constraining one or more physical parameters within the system, denoted ζ. A set of independent MD trajectories is then run, with different fixed values of the constraint(s) during each trajectory. Maintaining the constraints requires a force, f_ζ, which fluctuates throughout each trajectory. If the constraint is a single interatomic distance, the force is given by

$$f_\zeta = < \lambda >_\zeta \tag{4.1}$$

i.e. the ensemble average of λ, the Lagrange multiplier maintaining the constraint. The maximum and minimum values of the constrained distance are labelled ζ_1 and ζ_2. Sprik and Ciccotti showed that the Helmholtz free energy difference between the states ζ_1 and ζ_2 is given by

$$\Delta F = -\int_{\zeta 1}^{\zeta 2} f'_\zeta \, \mathrm{d}'_\zeta \tag{4.2}$$

i.e. by integrating over the force between the limiting values of ζ. This procedure is called thermodynamic integration.

Van Mourik and Gaigeot chose as their constraint the N3-H3 distance in BrU and U. For each of the two systems (BrU or U with 49 water molecules), six values of the constraint were applied, yielding 12 independent trajectories. For BrU, the constrained distances were 1.04, 1.23, 1.37, 1.60, 1.88 and 2.17 Å. For U, they were 1.05, 1.21, 1.34, 1.48, 1.61 and 1.94 Å. The longest values of this distance corresponded to starting structures in which the base had been deprotonated at N3, and a hydronium ion had been created by bringing the H3 proton to within bonding distance of the nearest water molecule. For BrU, when N3-H3 = 2.17 Å, the starting value of H3---O_w was 1.10 Å, while for U, when N3-H3 = 1.94 Å, the starting value of H3---O_w was 0.94 Å. Thermodynamic integration would therefore provide the Helmholtz free energy of aqueous deprotonation of the diketo tautomers of each base at N3, i.e. the following reaction:

$$\text{A-H3} + \text{H}_2\text{O} \rightarrow \text{A}^- + [\text{H3-H}_2\text{O}]^+ \tag{4.3}$$

where A = uracyl or 5-bromouracyl, and $[H3\text{-}H_2O]^+$ is the hydronium ion product.

Qualitatively, the results were as follows. For both bases, in the simulations with the longest N3-H3 constraints (2.17 or 1.94 Å), one of the unconstrained protons in the hydronium dissociated from H_3O^+ and transferred to a nearby water molecule, creating a new hydronium. In the case of uracil, the freed proton then transferred through a series of three further water molecules ("proton hopping"), remaining in solution throughout the trajectory. Hence the base was deprotonated and the water network was protonated, giving separated ions. For BrU, similar proton hopping occurred in the trajectories with the second- and third-longest N3-H3 constraints (1.60 and 1.88 Å). However, with the longest constraint (2.17 Å), the mobile proton first transferred to a nearby water molecule, then to the O4 atom of BrU. Here it remained, forming the O4-enol tautomer. Hence for BrU, but not U, proton transfer mediated by a bimolecular water bridge resulted in tautomerisation to the "mutagenic" form.

Note that this stepwise, trimolecular process, involving BrU and two bridging water molecules, differed from the tautomerisation of cytosine observed by Fogarasi[156] – that was a concerted, bimolecular process, in which an amino proton transferred to the single water molecule, while another proton simultaneously transferred from water to N3. Also note that the diffusion of the hydronium into solution, as observed in one U trajectory and two BrU trajectories, resembles the behaviour previously observed by Ivanov and Klein in CPMD simulations of the deprotonation of histidine, when their ζ value (also an N-H constraint) was larger than 1.6 Å.[160]

Van Mourik and Gaigeot also ran unconstrained simulations of the diketo tautomers of $U(H_2O)_{49}$ and $BrU(H_2O)_{49}$. The mean N3-H3 distances in these trajectories were 1.05 Å for $U(H_2O)_{49}$ and 1.04 Å for $BrU(H_2O)_{49}$.

4.2. N3-H3 constraints: thermodynamic integration

Using the data from van Mourik and Gaigeot, our first task was to analyse the Lagrange multipliers, λ, i.e. the forces maintaining the distance constraint throughout each of the 12 trajectories. These would lead to the free energies of deprotonation at N3.

Deprotonation was simulated *via* the creation of a hydronium ion near N3 in the setups with the longest N3-H3 constraints. However, as we have seen, in four trajectories this ion dissociated when a proton hopped into solution. It is crucial to realise that the values of λ, which maintain the constrained N3-H3 distance, only contain information about deprotonation for as long as the hydronium remains intact. Therefore, when calculating the mean force, $< \lambda >_\zeta$, for these trajectories, we should discard all values of λ taken at timesteps subsequent to dissociation. The timesteps at which this occurred, and the corresponding amounts of "real" simulation time prior to dissociation, were as follows:

BrU: ξ = 1.60 Å, timestep at H_3O^+ dissociation = 81180 ≡ 5.8 ps

BrU: ξ = 1.88 Å, timestep at H_3O^+ dissociation = 7159 ≡ 0.5 ps

BrU: ξ = 2.17 Å, timestep at H_3O^+ dissociation = 571 ≡ 0.04 ps

U: ξ = 1.94 Å, timestep at H_3O^+ dissociation = 2000 ≡ 0.2 ps

with timesteps of length 3.0 a.u. for BrU and 4.0 a.u. for U.

Bearing this in mind, we then calculated the relevant portions of $<\lambda>_\zeta$ for each trajectory, and plotted them against the corresponding constrained distances, ξ. These plots are shown in Figures 4.1 and 4.2, respectively, for U and BrU. For comparison, the graphs also show the values of $<\lambda>_\zeta$ calculated from *all* timesteps of each trajectory, even after dissociation of H_3O^+.

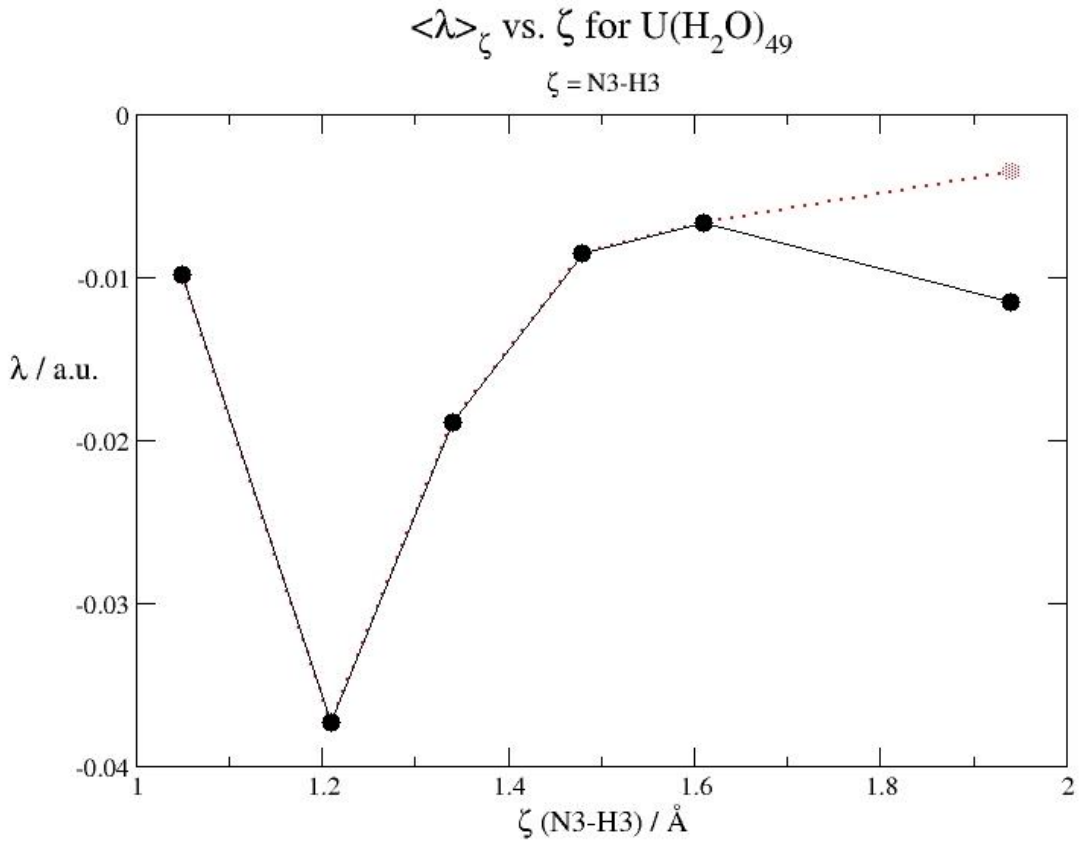


Figure 4.1: Mean force of constraint, $<\lambda>_\zeta$, as a function of the constrained N3-H3 distance, ζ, for uracil in a periodic water box. Calculated from raw data provided by van Mourik and Gaigeot.[161] The dotted red point shows the value of $<\lambda>_\zeta$ obtained for the largest value of ζ (1.94 Å) when the average is taken over all timesteps, including those subsequent to H_3O^+ dissociation. However, this point was discounted from the thermodynamic integration (see below).

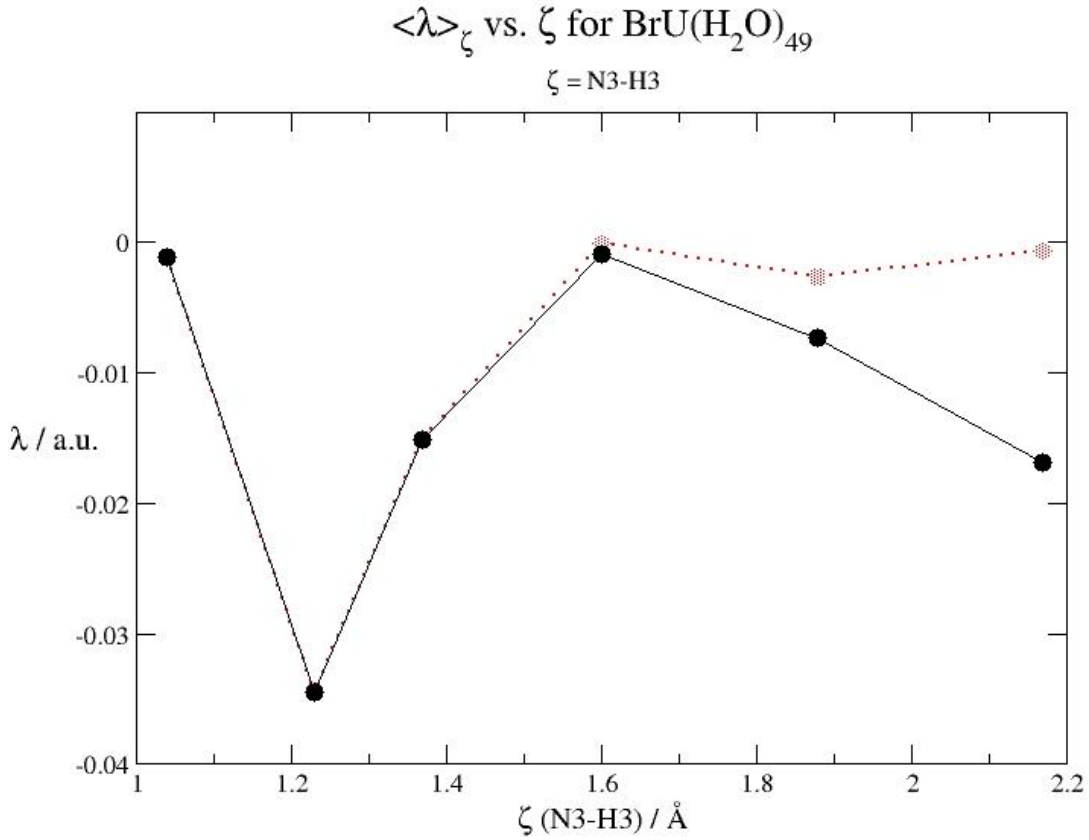


Figure 4.2: Mean force of constraint, $\langle\lambda\rangle_\zeta$, as a function of the constrained N3-H3 distance, ζ, for 5-bromouracil in a periodic water box. Calculated from raw data provided by van Mourik and Gaigeot.[161] The dotted red line shows the values of $\langle\lambda\rangle_\zeta$ obtained for the largest three values of ζ when the averages are taken over all timesteps, including those subsequent to H_3O^+ dissociation. However, the last two points were discounted from the thermodynamic integration (see below).

For both systems, the mean value of λ is near zero when the shortest N3-H3 constraint (≈1.05 Å in each case) is imposed. This state corresponds to the neutral diketo tautomer of each base, featuring a covalent N3-H3 bond which does not require much force to maintain. With the second-shortest value of ζ (≈1.2 Å for both systems), the force of constraint reaches its most negative value, indicating an intermediate state between diketo tautomer and contact ion pair. The force then increases with ζ, and comes close to zero again when the constrained distance is around 1.6 Å. This state corresponds to deprotonation at N3, forming a contact ion pair. Beyond this value of ζ, the mean $\langle \lambda \rangle_\zeta$ becomes strongly negative again: apparently, the longest N3-H3 constraints (greater than 1.8 Å) enforced geometries in which the hydronium was kept artificially far from the base, requiring some force to maintain.

We therefore imposed a further restriction on our thermodynamic integration: we would not integrate under the curves beyond the points at which $\langle \lambda \rangle_\zeta$ starts to drop from zero again. This meant discounting the last point (ζ = 1.94 Å) for uracil, and the last two points (ζ = 1.88 and 2.17 Å) for 5-bromouracil. Those are the three trajectories in which H_3O^+ dissociated in less than 1 ps.

We now used Grace[162] to integrate beneath the resulting curves (using five data points for U, and four for BrU). This integration took the values of ζ in bohr, rather than angstroms, and thus returned the

free energy in atomic units. The resulting free energies, converted to kcal/mol, are shown in Table 4.1.

System	$\Delta F_{\text{deprot(N3)}}$, kcal/mol
$U(H_2O)_{49}$	12.3
$BrU(H_2O)_{49}$	10.3

Table 4.1: Helmholtz free energy changes, ΔF, for deprotonation of uracil and 5-bromouracil at N3, in a periodic water box (Equation 4.3). Calculated from data provided by van Mourik and Gaigeot.[89]

The importance of discounting the points beyond which $<\lambda>_\zeta$ drops from zero again is illustrated in Figure 4.3, which shows the evolution of ΔF with ζ for $BrU(H_2O)_{49}$. The two largest values of ζ impose artificially long non-covalent distances between hydronium and N3, resulting in strongly negative values of $<\lambda>_\zeta$ (cf. Figure 4.2). This in turn leads to a large increase in ΔF, causing a gross overestimation of the penalty for deprotonation at N3 if these points – which represent regions of phase space that are unlikely to be sampled – are included in the integration.

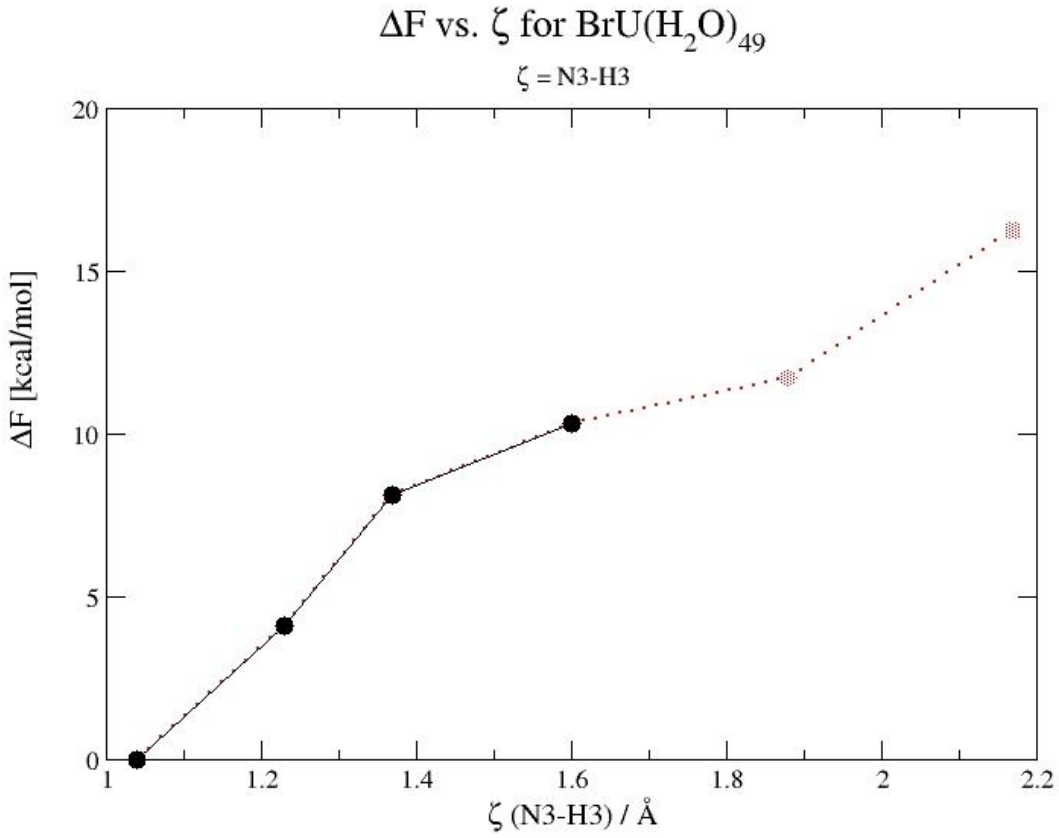


Figure 4.3: Evolution of the free energy, ΔF, with the distance constraint, ζ (N3-H3), for $BrU(H_2O)_{49}$. Each successive point is the result of integrating beneath the corresponding number of black points in Figure 4.2. The dotted red line indicates integration beneath the last two black points in Figure 4.2, which were in fact excluded from the final value of ΔF.

4.3. O4-H4 constraints: dynamics

The water-mediated tautomerisation of BrU observed in the trajectory with ζ = 2.17 Å was a three-step process, with the proton transfer steps being: (1) from N3 to the nearest water molecule, (2) from this water molecule to a neighbour, and (3) from the neighbour to O4. With step (1) modelled using distance constraints, we have calculated the free energies of deprotonation for both U and BrU, using van Mourik and Gaigeot's data. Assuming that step (2) is relatively facile, a complete thermodynamic picture of tautomerisation then requires calculating the free energy of step (3) – protonation at O4. This can be approximated as the negative of the energy of the reverse process, deprotonation at O4, i.e. the following reaction between the O4-enol tautomer and water:

$$\text{A-H4} + H_2O \rightarrow A^- + [\text{H4-}H_2O]^+ \qquad (4.4)$$

where A = uracyl or 5-bromouracyl, and $[\text{H4-}H_2O]^+$ is the hydronium ion product. This deprotonation has, in principle, the same endpoint as Equation 4.3, i.e. a contact ion pair between the deprotonated enolate and hydronium. In practice, there is probably a free energy difference between the two resulting species, owing to the different ionic interaction sites (N3 and O4). However, this error is likely to be partially offset by the fact that the free energies of separation of the two contact ion pairs – which presumably also differ – are already neglected in the free energy calculations.

To calculate the free energy of the deprotonation in Equation 4.4, we ran a further set of 12 dynamical trajectories (six for each base, with 49 water molecules, as before). In each starting structure, the base was already fully deprotonated at N3, by simply deleting H3. A new proton (labelled H4) was then inserted between O4 and the nearest water oxygen (O_w), forming a linear structure, O4-H4-O_w, with increasing values of the O4-H4 distance. The chosen values of this distance for BrU were 1.00, 1.14, 1.28, 1.42, 1.56 and 1.71 Å, while for U, they were 1.06, 1.22, 1.38, 1.54, 1.70 and 1.86 Å. The increasing O4-H4 distances were accompanied by decreasing H4-O_w distances: for BrU, the maximum value of H4-O_w was 1.75 Å (when O4-H4 = 1.00 Å), while the minimum was 1.04 Å (when O4-H4 = 1.71 Å). Likewise, for U, H4-O_w varied between 1.67 and 0.96 Å. Therefore, the smallest values of O4-H4 corresponded to the O4-enol tautomer interacting non-covalently with water at O4, while the largest values corresponded to the deprotonated enolate interacting non-covalently with hydronium at O4 (i.e. a contact ion pair). Figures 4.4 and 4.5 show the starting geometries for BrU with O4-H4 = 1.00 and 1.71 Å, respectively.

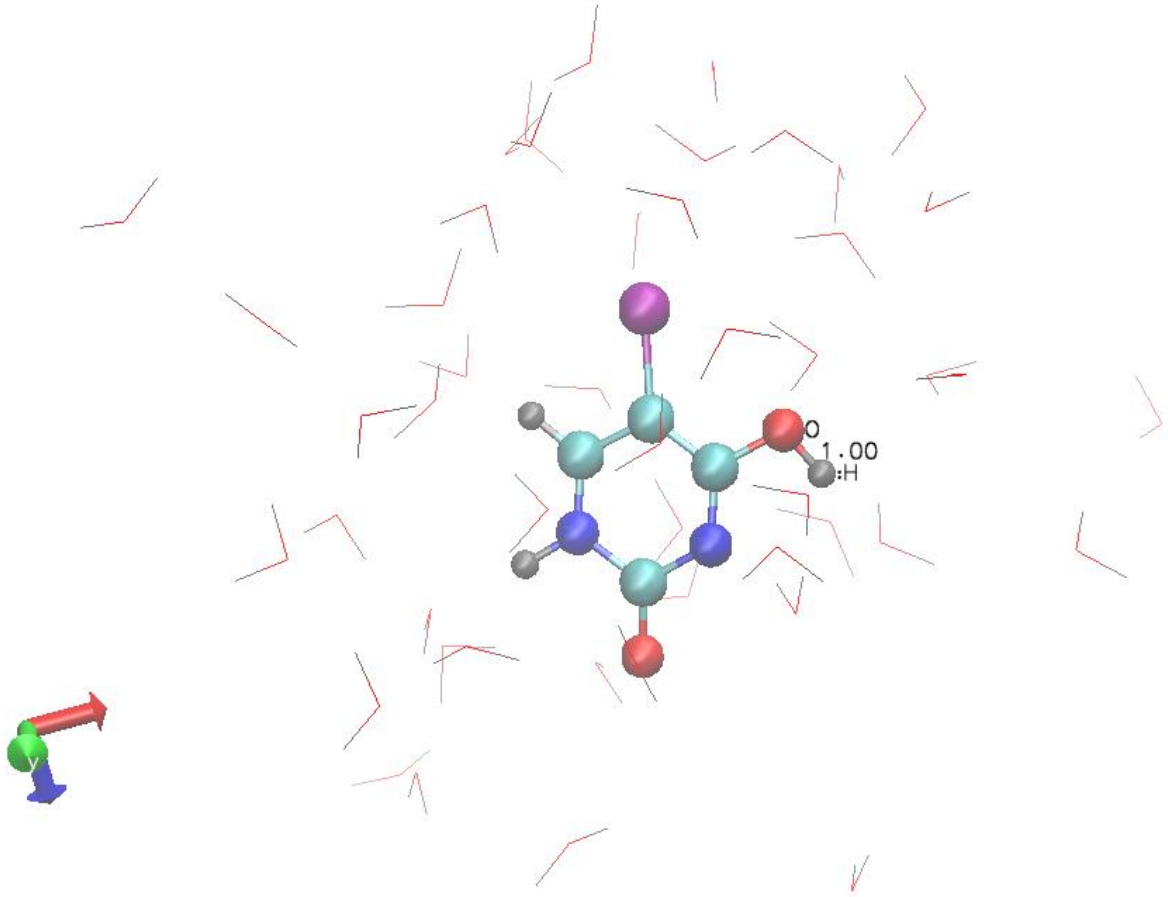


Figure 4.4: Starting geometry of the $BrU(H_2O)_{49}$ trajectory with ζ (O4-H4) = 1.00 Å (distance labelled).

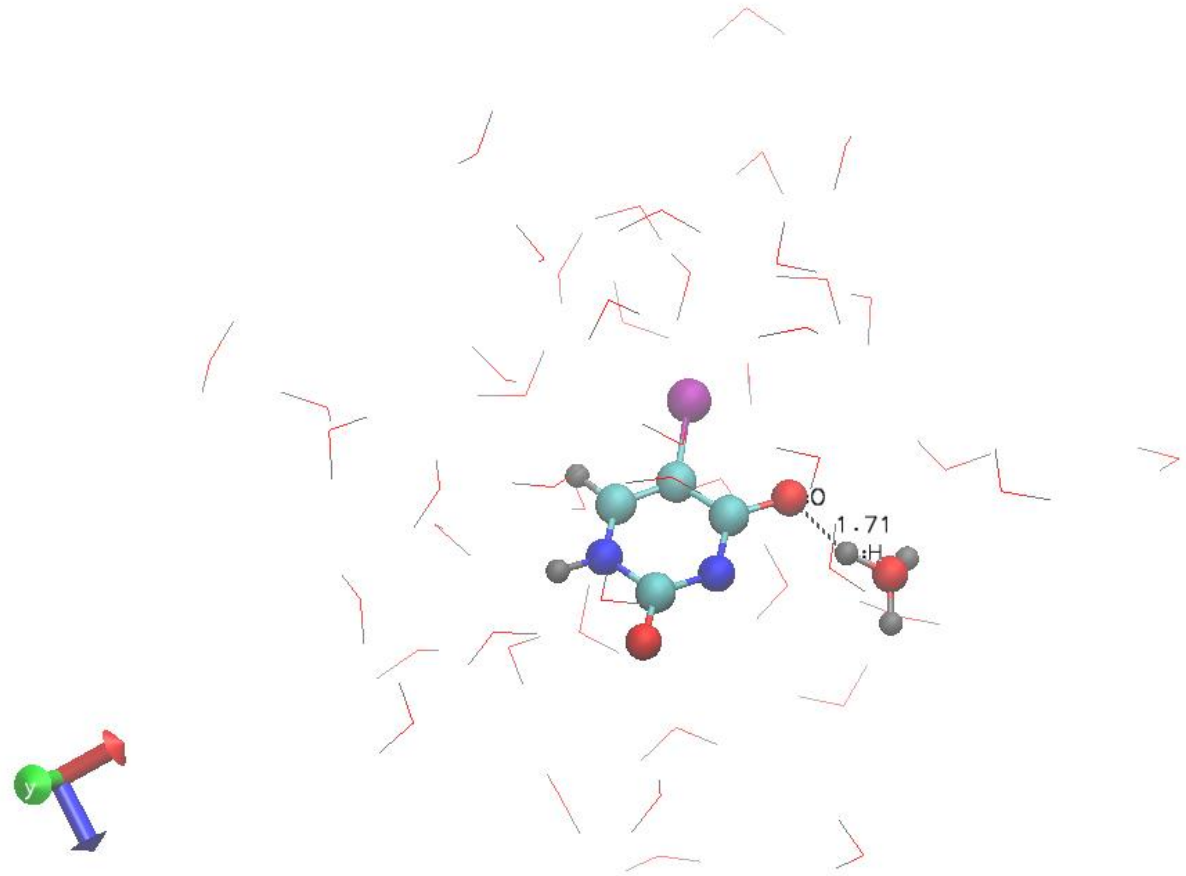


Figure 4.5: Starting geometry of the $BrU(H_2O)_{49}$ trajectory with ζ (O4-H4) = 1.71 Å (distance labelled). The resulting hydronium molecule is also highlighted.

In each trajectory, the O4-H4 distance (henceforth labelled ζ) was constrained to its initial value throughout the simulation.

These trajectories followed the general setup of van Mourik and Gaigeot's previous simulations. All simulations were carried out with the CPMD package.[163] For the exchange and correlation terms, we used BLYP-D, i.e. the Becke–Lee–Yang–Parr GGA functional[110, 113, 164] with Grimme D2-type dispersion correction.[117] The one-electron orbitals were expanded in a plane wave basis set with a kinetic energy cut-off of 90 Ry restricted to the gamma point of the Brillouin zone. Medium-soft norm-conserving pseudopotentials of the Martins–Troullier type[165] were used. The core–valence interaction in C, N, and O was treated by s and p potentials with pseudisation radii of 1.23, 1.12, and 1.05 a.u., respectively (taking the same radius for s and p), whereas H atoms were treated as an s potential with a radius of 0.5 a.u. Energy expectation values were calculated in reciprocal space using the Kleinman–Bylander transformation.[166] Dynamics were performed in the microcanonical ensemble (constant volume and internal energy) using a fictitious electron mass of 400 a.u. (for U) or 300 a.u. (for BrU) and a timestep of 4.0 a.u. (for U) or 3.0 a.u. (for BrU). Periodic boundary conditions were applied to a cubic box length of 11.5 Å.

The volume and the number of molecules in the box were those used by van Mourik and Gaigeot in their previous study, and had been chosen by them in order to provide a density of liquid water of 1 g/cm^3 together with sufficient layers of water surrounding the solute. The initial conformations of all

the trajectories (modified per the constraints) were also based on the previous investigation by those authors.

Each trajectory was run for 10 ps of simulation time, split into three phases: equilibration (1 ps), thermalisation (3 ps) and production (6 ps). During equilibration, the temperature was constrained to 320 K ± 50 K, by rescaling the velocities when the temperature went outside this range. Rescaling was not performed during thermalisation or production.

We then used VMD[167, 168] to view the resulting dynamics. For BrU, dissociation of H_3O^+ occurred with two values of ζ: 1.42 Å (the third largest) and 1.71 Å (the largest). For U, dissociation occurred with the two largest values of ζ: 1.70 Å and 1.86 Å. The timesteps at which this occurred, and the corresponding amounts of "real" simulation time prior to dissociation, were as follows:

BrU: ξ = 1.42 Å, timestep at H_3O^+ dissociation = 114175 ≡ 8.2 ps

BrU: ξ = 1.71 Å, timestep at H_3O^+ dissociation = 3450 ≡ 0.2 ps

U: ξ = 1.70 Å, timestep at H_3O^+ dissociation = 298 ≡ 0.03 ps

U: ξ = 1.86 Å, timestep at H_3O^+ dissociation = 62 ≡ 0.01 ps

with timesteps of length 3.0 a.u. for BrU and 4.0 a.u. for U.

Reprotonation of the bases did not occur in any trajectory.

We also ran unconstrained simulations of the O4-enol tautomers of $U(H_2O)_{49}$ and $BrU(H_2O)_{49}$. The mean O4-H4 distances in these trajectories were 1.02 Å for $U(H_2O)_{49}$ and 1.04 Å for $BrU(H_2O)_{49}$.

4.4. O4-H4 constraints: thermodynamic integration

The initial water networks required periods of equilibration and thermalisation, to accommodate to the new geometries of the bases with their distance constraints. It is therefore important to check that the constraining force, λ, had reached a reasonably steady value by the beginning of the production phase (6 ps) in each trajectory. We calculated the running average of λ over each trajectory, using an in-house Fortran 90 program written by Prof. Michael Bühl.[169] In each case, $< \lambda >_\zeta$ had become stable within 6 ps. An example of these running averages – for $BrU(H_2O)_{49}$ with ζ (O4-H4) = 1.56 Å – is plotted in Figure 4.6.

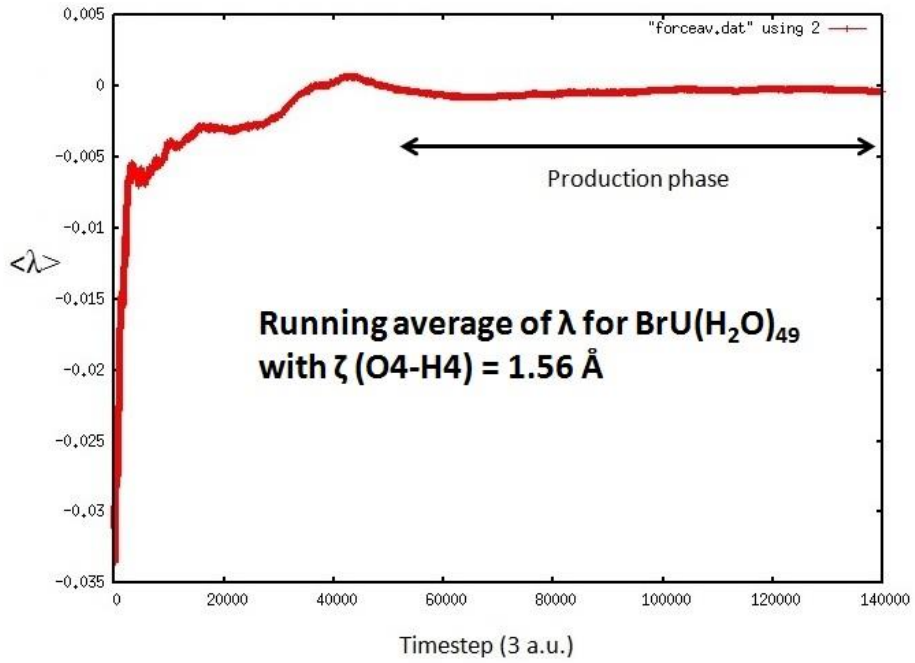


Figure 4.6: Running average of the force, λ, during the $BrU(H_2O)_{49}$ trajectory with ζ (O4-H4) = 1.56 Å.

As before, we then needed to calculate the final values of $< \lambda >_\zeta$ for each trajectory, and plot them against ξ. For three trajectories (the last three listed in the previous paragraph), H_3O^+ dissociation occurred during the equilibration phase (i.e. the first picosecond of simulation time). In these cases, we calculated $< \lambda >_\zeta$ from the *start* of the simulation, until dissociation. For the other nine trajectories, we restricted $< \lambda >_\zeta$ to the timesteps in the production phase (after 6 ps). For the BrU trajectory with ξ = 1.42 Å, the timesteps subsequent to H_3O^+ dissociation (8.2 ps) were discarded. The resulting plots of $< \lambda >_\zeta$ vs. ζ are shown in Figures 4.7 and 4.8, for U and BrU, respectively. For comparison, the graphs also show the values of $< \lambda >_\zeta$ calculated over all the production timesteps of each trajectory.

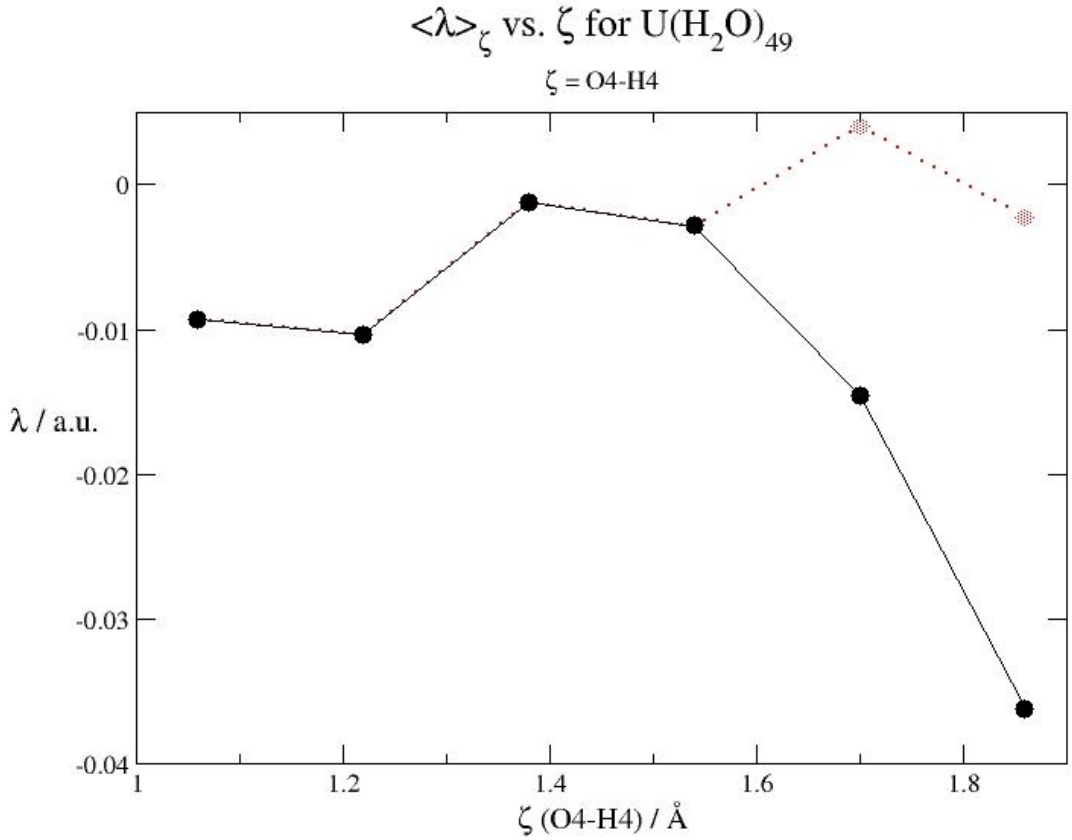


Figure 4.7: Mean force of constraint, $\langle\lambda\rangle_\zeta$, as a function of the constrained O4-H4 distance, ζ, for uracil in a periodic water box. The dotted red line shows the values of $\langle\lambda\rangle_\zeta$ obtained for the largest two values of ζ when the averages are taken over all timesteps, including those subsequent to H_3O^+ dissociation. However, these points were discounted from the thermodynamic integration (see below).

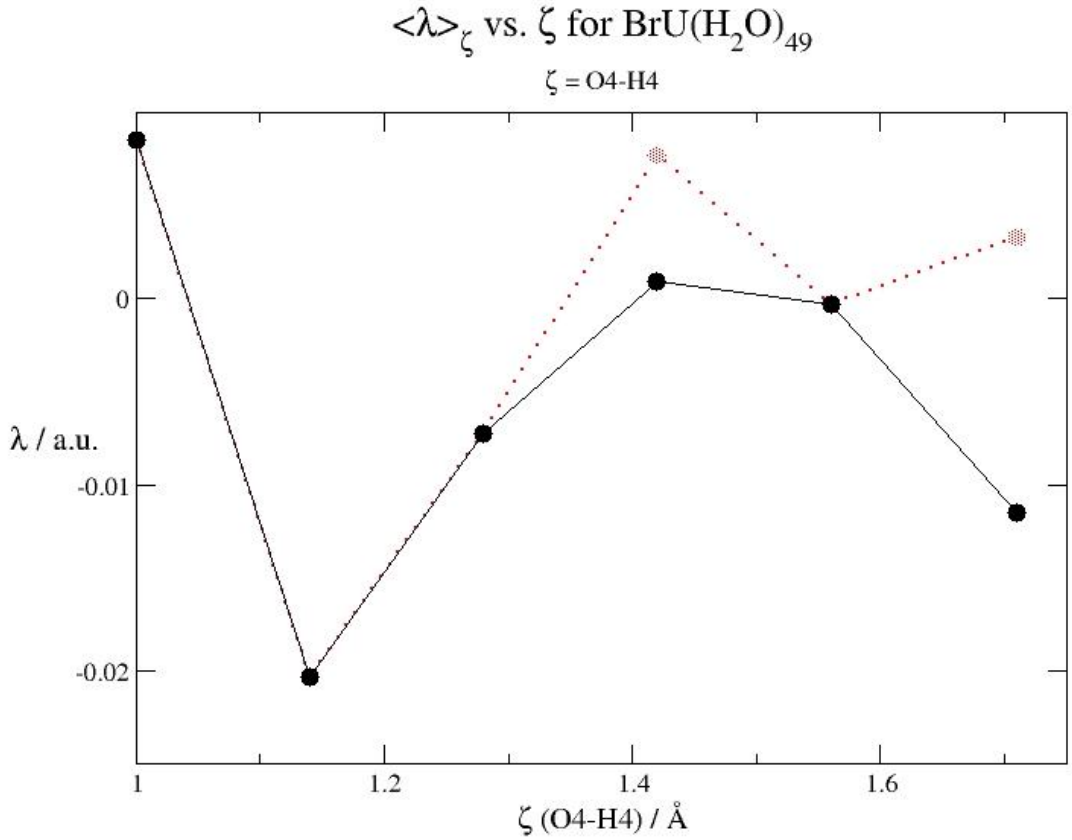


Figure 4.8: Mean force of constraint, $\langle\lambda\rangle_\zeta$, as a function of the constrained O4-H4 distance, ζ, for 5-bromouracil in a periodic water box. The dotted red line shows the values of $\langle\lambda\rangle_\zeta$ obtained for the third-largest and largest values of ζ when the averages are taken over all timesteps, including those subsequent to H_3O^+ dissociation. However, the final point was discounted from the thermodynamic integration (see below).

As with the N3-H3 constraint, it can be seen that, for both systems, $< \lambda >_\zeta$ first reaches a minimum value, then approximately reaches zero, then becomes more negative again with the largest values of ζ. For U, the first minimum of the force is at ζ = 1.22 Å, while for BrU, it is at ζ = 1.14 Å. These minima correspond to a situation intermediate between O4-enol and contact ion pair, while the values close to zero (when ζ ≈ 1.4 Å) correspond to the contact ion pairs. Beyond those points, it seems that ζ is too large, as with N3-H3. Therefore, on the basis of these plots, we decided to discount the two largest values of ζ (1.70 and 1.86 Å) from the thermodynamic integration for U, while for BrU, we discounted the largest value (1.71 Å). This means that the three trajectories in which H_3O^+ dissociated during the equilibration phase were discounted from the final free energy calculation.

To summarise: for U, thermodynamic integration was performed using the mean values of λ, evaluated after 6 ps, for the trajectories with the four lowest values of ζ; while for BrU, we used the mean λ after 6 ps for the trajectories with the *five* lowest values of ζ, but excluding the timesteps after 8.2 ps for ζ = 1.42 Å. The resulting free energies are shown in Table 4.2.

System	$\Delta F_{\text{deprot(O4)}}$, kcal/mol
$U(H_2O)_{49}$	3.4
$BrU(H_2O)_{49}$	3.7

Table 4.2: Helmholtz free energy changes, Δ*F*, for deprotonation of uracil and 5-bromouracil at O4 in a periodic water box (Equation 4.4).

4.5. Discussion

Comparison of Tables 4.1 and 4.2 shows that, for both bases, the diketo form is more stable against deprotonation than the O4-enol form. Assuming that the Helmholtz free energies can be equated with the Gibbs free energies, Δ*G* (i.e. assuming that volume does not vary significantly with pressure), we can use Table 4.1 to estimate the pK_a of acidity of each base at N3, according to:[160]

$$\Delta G = 2.303 RT \mathrm{p}K_a \tag{4.5}$$

With *T* = 320 K, solving for the right-hand side yields pK_a(N3) = **8.4** for uracil and **7.1** for 5-bromouracil: BrU is a stronger proton donor. In 1962, Katritzky and Waring estimated these values as 9.5 and 7.8, respectively, by titration at 297 K.[65] More recent experimental estimates for uracil have been 9.3 and 9.8,[170, 171] while B3LYP/continuum-solvent calculations by Jang *et al.* yielded a value for uracil at 298 K of 9.3.[172] In 1989, Sowers *et al.* determined that the pK_a of the dBrU nucleoside at 296 K was 8.1.[83]

The results in Table 4.2 can likewise be used to calculate the pK_a of each base at O4: solving Equation 4.5 yields pK_a(O4) = **2.3** for uracil and **2.5** for 5-bromouracil.

The difference between the two pK_a's for each base can be used to estimate the aqueous free energies of tautomerisation, $\Delta F_{\text{keto}\rightarrow\text{enol}}$ (aq.), by setting pK_a = pK_a(N3) − pK_a(O4), then solving for the left-hand side of Equation 4.5 This amounts to taking the difference between the deprotonation free energies at the two sites.[79] The tautomerisation energies estimated in this way are markedly different for the two bases: for uracil, $\Delta F_{\text{keto}\rightarrow\text{enol}}$ (aq.) ≈ 8.9 kcal/mol, while for 5-bromouracil, $\Delta F_{\text{keto}\rightarrow\text{enol}}$ (aq.) ≈ 6.6 kcal/mol. Tautomerisation is therefore more favourable for BrU than for U by 2.3 kcal/mol. Recalling Loeb and Kunkel's estimation[19] that a free energy difference of 1.4*n* kcal/mol should cause a mispairing rate of 1 in 10^n, this implies that BrU should mispair more often than U by a factor of $10^{2.3\div1.4)}$ = 44. (Here we neglect any changes in pK_a upon going from the free nucleobases, U and BrU, to the bound nucleotides in a polymer.)

The difference between the aqueous free energies of tautomerisation derived here, and the corresponding gas-phase potential energies given in Chapter 3 (Table 3.1), quantifies the energetic effect of solvation on the tautomerism of U and BrU. This term, which is made up of enthalpic,

entropic and solute–solvent contributions, can then be used to estimate the aqueous tautomerisation energies at higher levels of theory, for which only the gas-phase term is known, according to:

$$\Delta F^{\text{high-level}}_{\text{keto}\rightarrow\text{enol}}\,(\text{aq.}) \approx \Delta E^{\text{high-level}}_{\text{keto}\rightarrow\text{enol}}\,(\text{gas}) + \left[\Delta F^{\text{BLYP}}_{\text{keto}\rightarrow\text{enol}}\,(\text{aq.}) - \Delta E^{\text{BLYP}}_{\text{keto}\rightarrow\text{enol}}\,(\text{gas})\right] \quad (4.6)$$

Here we assume that the effect of solvation (given in square brackets) is method-invariant, so the BLYP value can be applied to higher levels of theory. The standard, all-electron BLYP/6-31G(d,p) values of $\Delta E_{\text{keto}\rightarrow\text{enol}}$ (gas) calculated with Gaussian 09 are 12.1 kcal/mol for U and 12.3 kcal/mol for BrU (Chapter 3, Table 3.1). However, the BLYP/plane-wave values, calculated by us using CPMD, were found to be 11.2 and 12.0 kcal/mol. Thus the contribution of solvation to the tautomerisation free energy (derived from the $\Delta F^{\text{BLYP}}_{\text{keto}\rightarrow\text{enol}}\,(\text{aq.})$ values of 8.9 kcal/mol for U and 6.6 kcal/mol for BrU) is −2.3 kcal/mol for U and −5.4 kcal/mol for BrU, when using the BLYP/plane-wave gas-phase energies. Applying these corrections to the gas-phase M06-2X/6-31G(d,p) energies (Chapter 3, Table 3.1) yields $\Delta F^{\text{M06-2X/6-31G(d,p)}}_{\text{keto}\rightarrow\text{enol}}\,(\text{aq.})$ = 8.6 kcal/mol for U and 6.0 kcal/mol for BrU. The difference in tautomerisation energies is therefore 2.6 kcal/mol at this level of theory, which by Loeb and Kunkel's equation implies a BrU-to-U mispairing ratio of $10^{(2.6\div1.4)}$ = 72, compared to 44 for BLYP/CPMD (given on the previous page).

Alternatively, using the all-electron BLYP values of $\Delta E_{\text{keto}\rightarrow\text{enol}}$ (gas), the contribution of solvation to the free energies is −3.2 kcal/mol for U and −5.7 kcal/mol for BrU. This yields M06-2X/6-31G(d,p) tautomerisation free energies of 7.6 kcal/mol for U and 5.7 kcal/mol for BrU, implying a BrU-to-U mispairing ratio of $10^{(1.9\div1.4)}$ = 24, compared to 44 for BLYP/CPMD. Therefore, solvation-corrected M06-2X predicts either a higher or lower mispairing ratio of BrU to U than BLYP/CPMD itself does, depending on which gas-phase BLYP values are used to quantify the effect of solvation, but in any case the ratios are all of the same order of magnitude.

4.5.1. Error evaluation

Various sources of error in the final results need to be considered.

Firstly, the use of a *single* distance constraint in each set of trajectories may have imposed some bias. Here we have run two separate series of constraints (ζ = N3-H3 or O4-H4) for each base, generating separate free energy measurements of the two protonation/deprotonation steps. The energetics of proton transfer from N3 to O4 – a multi-step process, potentially involving several H atoms – cannot be modelled by a unique sequence of trajectories if the same single distance constraint is employed in each of them.[173] Such a seamless set of MD trajectories would require a more sophisticated choice of reaction coordinate. For example, Sprik[174] modelled the auto-dissociation of water using a coordination number constraint, constraining the number of protons within bonding distance of a chosen oxygen. By applying a weighting function to protons at intermediate O-H distances, the constraint could be varied smoothly towards either 3 (enforcing the creation of a hydronium) or 1 (enforcing the creation of a hydroxyl). This in principle provides a more flexible description of the reaction coordinate than a simple distance constraint. However, the final energy was found to be independent of the type of constraint used. Another choice is the "difference-of-distances", in which a triatomic bond-breaking/bond-making process, A-B + C → A + B-C, is studied by constraining the difference between the A-B and B-C distances.[175]

Secondly, uncertainties in the force of constraint, λ, lead to a statistical error, $\delta\Delta F$, in each free energy. $\delta\Delta F$ can be estimated by measuring the standard deviation of the running average of λ during

the last picosecond of each trajectory, then multiplying the largest of these by the total integration width.[169] In the N3-H3 trajectories, the largest standard deviations (measured using an in-house program provided by Prof. Michael Bühl) were 3.9 x10^{-4} a.u. for uracil (when ζ = 1.34 Å), and 1.3 x 10^{-3} a.u. for 5-bromouracil (when ζ = 1.04 Å). The integration widths were 1.06 and 1.05 bohr, respectively. This yields $\delta\Delta F$ = ±**0.26 kcal/mol** for uracil and ±**0.87 kcal/mol** for 5-bromouracil. In the O4-H4 trajectories, the largest standard deviations were 2.4 x10^{-4} a.u. for uracil (when ζ = 1.06 Å), and 2.0 x 10^{-3} a.u. for 5-bromouracil (when ζ = 1.42 Å). The integration widths were 0.91 and 1.06 bohr, respectively. Hence $\delta\Delta F$ = ±**0.14 kcal/mol** for uracil and ±**1.33 kcal/mol** for 5-bromouracil. A more sophisticated method of estimating $\delta\Delta F$ from the uncertainty in λ, based on Simpson's rule, was developed by Senn *et al*.,[175] but since it is probably not the largest source of numerical error in the first place, further precision in its estimation is not necessary.

Thirdly, the choice of functional plays a role. The use of BLYP for simulations of small biomolecules immersed in liquid water is well established in the group of Gaigeot.[173, 176] However, other authors have detected weaknesses in the dynamical description of water by this method. In CPMD simulations of the transport of a hydrated excess proton, Izvekov and Voth noted that BLYP caused an unphysical over-structuring of water, resulting in an attraction between the excess proton and one "special" water oxygen, and thus significantly reducing the self-diffusion coefficient of water compared to experiment.[177] Likewise, Todorova *et al.* observed over-structuring of water by BLYP, and recommended hybrid functionals for the more accurate calculation of self-diffusion coefficients and radial distribution functions.[178] However, our study was not concerned with the overall water structure, nor with the behaviour of the hydronium proton once it had diffused into solution, but only with the constrained base–water interactions at N3 and O4.

Fourthly, the use of only four or five integration points per free energy calculation represents a fairly coarse mesh, which introduces another source of numerical error. Six points per energy would have been available had all the trajectories yielded useful data, but it became evident that the longest values of ζ we had chosen were too long – and even six is still relatively few. It would now be desirable to perform further simulations, in which ζ is constrained to values slightly above and slightly below those generating the maximum force (around 1.2 Å for all four systems). These extra integration points would create a finer mesh in the regions of the graphs that contribute most to the final values of ΔF. In the longer term, it would also be valuable to repeat these sets of simulations using thymine as the base, in order to quantify the effect of the 5-methyl substituent on the pK_a of T compared to U.

Our overall conclusion from this study is that the 5-bromine substituent lowers the pK_a of BrU compared to uracil, potentially increasing the likelihood that the "mutagenic" enol or ionic tautomers are formed in water.

Chapter 5: Base stacking and mutagenicity

5.1. Background

Vertical (intra-strand) base stacking is a crucial component of DNA structure. Perturbation of stacking is therefore expected to cause structural changes that could interfere with the normal function of DNA. This chapter focuses on the possibility that enhanced base stacking of BrU is a determinant of its mutagenicity.

First, a general description of the geometry and energetics of stacking is given. Previous studies on the stacking of BrU are then reviewed. Finally, our own results are presented.

5.1.1. Geometric description of base stacking

Stacking of base pairs can be quantified using a set of geometric descriptors, known as base step parameters, which describe the relative orientation of two base pairs in a step. These parameters are defined using a Cartesian coordinate frame which is independent of the global helical axis (Figure 5.1). If the two strands are arbitrarily designated strand I and strand II, then the *y*-axis for each base pair points from strand II to strand I along the C6(pyrimidine)–C8(purine) vector in the plane of the base pair. The midpoint between those two atoms is the axis origin. The *z*-axis is normal to the base pair plane, and points in the 5' → 3' direction of strand I. The *x*-axis completes a right-hand set, being normal to *y* in the base pair plane.[179-181]

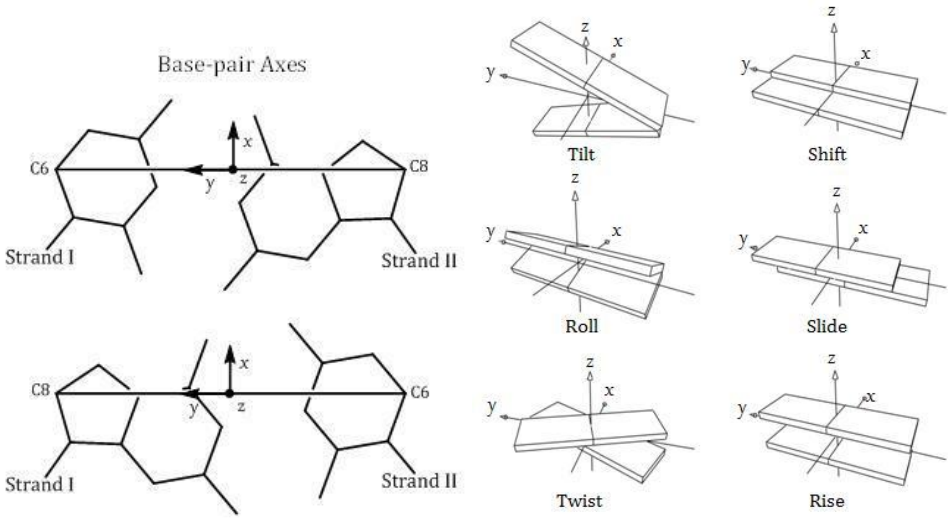


Figure 5.1: Definition of the base step parameters. Left: coordinate frame. Centre: rotations. Right: translations. Modified from an image provided by Dr. Darren Smith, based on Refs [173] and [175].

Each base step parameter is then either a rotation of one base pair about an axis, or a translation of one base pair along an axis, relative to the other base pair (Figure 5.1). The six parameters are Tilt (rotation about the *x* axis), Roll (rotation about *y*), Twist (rotation about *z*), Shift (translation along *x*), Slide (translation along *y*) and Rise (translation along *z*).[179-181]

Table 5.1 shows the average values of these quantities in conventional ("B-form") DNA, obtained from crystal structures by Olson *et al.*[182] Two base step parameters are by necessity non-zero in DNA: Twist, which results in the classic helical shape of B-form DNA, and is typically ~36°, and

Rise, which results from short-range inter-base repulsion, and is typically 3.3 or 3.4 Å.[182, 183] Tilt, Roll, Shift and Slide are zero in an idealised B-DNA helix, and are usually close to zero in practice.

Base step rotations (angles)			Base step translations (distances)		
Tilt (°)	Roll (°)	Twist (°)	Shift (Å)	Slide (Å)	Rise (Å)
−0.1	**0.6**	**36.0**	**−0.02**	**0.23**	**3.32**

Table 5.1. Average experimental values of base step parameters in B-form DNA, from Olson *et al.*[182]

5.1.2. Physical components of base stacking

The balance of energetic contributions to base stacking has been the focus of a considerable body of research over many decades. A comprehensive review of this vast, multidisciplinary field, which is primarily experimental, would be beyond the scope of a theoretical chemistry thesis focused on one base analogue. However, an overview of historical and contemporary research will be given, concentrating on recent theoretical results.

DNA is typically studied in aqueous solution. Its stability against denaturation (loss of secondary structure) is a function of temperature, pH and ionic strength, and is also reduced in non-aqueous solvents.[184] Although Watson and Crick initially focused on hydrogen bonding as the stabilising interaction,[1, 5] by the end of the 1950s it was already clear, from several lines of evidence, that this could not be the sole contribution.[184] In the 1960s it was shown that even isolated bases could form stacked polymers in aqueous solution, and the enhanced stacking of methylated and brominated base analogues was attributed to van der Waals forces.[185] Bugg *et al.* subsequently observed by X-ray diffraction that the stacking of pure crystalline DNA monomers, e.g. guanosine, could not be driven by the interaction of permanent dipoles, since the bases often stacked with parallel dipole moments; instead, they argued for the role of dipole–induced dipole forces.[77]

By the 1990s it was accepted that electrostatics, dispersion and the hydrophobic effect all played a role, but their relative importance had not been established.[11] In 1995, Friedman and Honig published a theoretical study of all possible stacked combinations of the natural bases.[11] Thermodynamic cycles and a continuum model of water were used to calculate the free energy of transferring isolated bases in the gas phase to stacked dimers in solution. The two contributions to the free energy (the non-polar and electrostatic terms) were calculated with potential functions. The total electrostatic term was always unfavourable – however, the Coulombic component of the electrostatic term was variably favourable or unfavourable, while the reaction field component of the electrostatics (associated with desolvation upon stacking) was usually unfavourable. In contrast, the Lennard-Jones and non-polar solvation terms were always favourable. The authors concluded that stacking was driven by non-polar interactions, including dispersion and the hydrophobic effect.

An experimental study by Gellman *et al.* challenged this finding.[186] They argued that the small chemical shift in a non-polar bis-naphthyl compound, compared to larger shifts in analogous polar compounds, was evidence against the existence of bis-naphthyl stacking, which would rely on non-polar dispersive interactions. They concluded that stacking was instead driven by electrostatics. However, Friedman and Honig countered that Gellman *et al.* had misinterpreted their NMR data by failing to note that polar compounds produce greater chemical shifts than non-polar compounds.[187] They reiterated that stacking was apparently driven by non-polar effects. This was supported by Luo *et al.*, who argued that the weakness of bis-naphthyl stacking in the study by Gellman *et al.* was probably caused by the greater separation distance and reduced steric accessibility of favourable

stacking motifs in the chosen naphthyl structures, rather than their non-polarity. They also produced *ab initio* results showing non-electrostatic forces to dominate in both the polar and non-polar species.[188]

In 1999, Florián *et al.* calculated contributions to the interaction energies of stacked dimers and base pairs, in the gas phase and solution, using MP2 and a dipole solvation model.[189] They also estimated the free energies, using an experimentally established relationship between enthalpy and entropy. They found a large compensation between gas-phase and solvation energetics: the energetic effects introduced by the solvent were most destabilising at geometries that were most favourable in the gas phase, leading to flat potential energy surfaces. Electron correlation and solvation were found to have a similar Twist dependence (favouring 0° Twist); methyl, amino and keto substituents enhanced stacking; and H-bond and stacking energies for a given combination of bases tended to be similar.

In 2000, Guckian *et al.* performed a systematic experimental attempt to determine the relative contributions of different forces to stacking.[12] They synthesized DNA heptamers containing a single "dangling nucleoside" – an unpaired base at the end of the strand, which takes part in stacking but not hydrogen bonding. The identity of the unpaired base was varied, to include all the natural bases, plus a selection of unnatural base analogues, featuring non-polar nucleoside mimics, aromatic hydrocarbons and electron-withdrawing groups. By thermal denaturation, they measured the thermodynamic stabilisation afforded by the stacking of each unpaired base. These free energy terms varied from −0.8 to −3.4 kcal/mol. The energies were plotted against calculated physical properties (hydrophobicity, dipole moment, polarisability and surface area). They found no correlation of hydrophobicity or dipole moment with stacking energy, but a qualitative correlation of polarisability and surface-area-excluded with stacking energy. The natural bases stacked in the order A > G ≥ T = C, with limiting values of −2.0 kcal/mol (for A) and −1.0 kcal/mol (for C). The authors determined that, although dispersion was always stabilising, it was not a major contributor to the total stacking energy, since naphthalene and phenanthrene had very different polarisabilities but similar stacking energies, while the reverse was true for thymine and difluorotoluene. They also downplayed the involvement of electrostatic (dipole–dipole) forces, since most of the strongest stackers had small permanent dipole moments. They concluded that the single most important driving force for stacking was the hydrophobic effect (the surfaces of planar aromatics interact unfavourably with water, so in aqueous solution they stack in order to shield their surfaces from the solvent), and this effect is proportional to surface area.

A review by Kool in 2001 was more cautious, noting that further work was needed to properly establish the size of the dispersive contribution.[13] The author also noted that electrostatics may be *locally* significant for determining the geometry of stacking, but that its global energetic contribution is on average roughly zero, due to cancelling of positive and negative terms. A typical stacking free energy for a single natural base was given as roughly −1.0 kcal/mol.

In 2003, Jurečka and Hobza calculated the gas-phase energies and enthalpies of both Watson–Crick base pairing and stacking of the A-T and G-C dimers, at the CCSD(T)/CBS//RI-MP2/TZVPP level of theory.[190] The enthalpies were in good agreement with gas-phase experiments. Although no direct insights into the physical origins of stacking were obtained, the study gave the unexpected result that the G/C stacked dimer had a greater enthalpy (−15.0 kcal/mol) than the A-T base pair (−14.0 kcal/mol). It had once been believed that hydrogen bonding dominated over stacking both in the gas phase and in the helix. This inference had been drawn from the greater thermal stability of DNA

molecules that are rich in G-C pairs, which contain three H-bonds, over those rich in A-T pairs, which contain only two.[18] However, Jurečka and Hobza's study showed that stacking could be highly competitive with hydrogen bonding in the gas phase, even without the hydrophobic bias towards stacking introduced by a polar solvent. Methodologically, perhaps the most important conclusion was that the ΔCCSD(T) correction to the MP2 energies (a term which is almost always unfavourable for stacking) was necessary to obtain the correct relative energies of stacking and H-bonding.

In 2005, the same group performed a follow-up study, in which the same methodology was applied to base-paired and stacked structures derived from experimental DNA crystal structures in the Protein Data Bank (PDB).[191] While the hydrogen bonding energies were roughly the same in the crystal structures as in the gas phase, the stacking energies were sharply reduced, going from −12 to −8 kcal/mol for A/T, and from −17 to −8 kcal/mol for G/C. (Note that these are potential energies, not enthalpies.) The authors concluded that stacking and base pairing should contribute roughly equally to the stabilisation of AT-rich sequences of DNA, while in GC-rich sequences, stacking should only contribute around one third of the stabilisation. However, neither solvent effects nor entropy were included.

In 2006, Yakovchuk *et al.* provided direct experimental evidence that stacking completely dominates over base pairing in the helix.[18] While earlier studies had measured the thermal stability of short DNA stretches containing dangling nucleosides, this group used gel electrophoresis (relative mobility measurements) to determine the free energy perturbations caused by "nicks" – strategically placed sites featuring a single missing base – in long (300 base pair) DNA fragments. They estimated that the stacking of a G-C base pair contributes between −1.02 and −1.83 kcal/mol of free energy to the stability of the helix (depending on temperature and ionic concentration), while the contribution from hydrogen bonding was merely −0.01 to −0.11 kcal/mol. For an A-T base pair, stacking was found to be somewhat weaker (−0.36 to −1.32 kcal/mol), while hydrogen bonding was actually *de*stabilising, by approximately +0.60 kcal/mol. They concluded that the temperature dependence of DNA stability was entirely due to the stacking term, while the greater stacking of G-C base pairs was responsible for 50% of the dependence of DNA stability on the relative amount of G-C and A-T.

Although that study provided no direct insights into the physical origins of stacking, it strongly emphasized the need for such an understanding to be developed, since stacking was revealed to be the essential basis of DNA stability. The idea that base pairing should be weaker than stacking does make intuitive sense, when one considers that, during replication, the two strands must be easily "unzipped" from one another while remaining intact themselves.

Also in 2006, Šponer *et al.*[192] calculated reference stacking energies for all 10 unique DNA base steps, using CCSD(T) extrapolated[193] to the complete basis set limit. Comparison with force-field results showed that the force-fields performed surprisingly well. This implies that the physical contributions to stacking are relatively straightforward in themselves – i.e. they can be well reproduced in terms of familiar force-field parameters, such as polarisability (for dispersion) and dipole moment (for electrostatics) – and that the difficulty in quantitatively describing helical stacking is caused by the high sensitivity of these forces to subtle geometric changes.

A decomposition of the energetic components of stacking by quantum methods, rather than classical, was performed by Czyżnikowska *et al.* in 2007.[194] They studied the energy of the gas-phase uracil face-to-back dimer as a function of Twist and Rise at the CP-corrected df-MP2/aQZ level. For the

entire range of Rise values (3.0 to 5.0 Å), the most favourable Twist was 180°. The global minimum was at Rise = 3.3 Å. A hybrid variational-perturbational scheme was used to decompose the energy into electrostatic, Heitler–London exchange, delocalisation and dispersion terms at the aDZ level. At the global minimum, dispersion (−12.3 kcal/mol) was twice as stabilizing as first-order electrostatics (−6.5 kcal/mol), while the exchange term was +10.4 kcal/mol. The electrostatic and dispersion terms were stabilising in all geometries. However, the electrostatic terms depended strongly on Twist, while dispersion did not. Therefore dispersion was crucial for overall stacking stability, while electrostatics determined the Twist.

However, a Twist of 180° (with the N1–C4 vectors of the two uracil or thymine monomers pointing in opposite directions) is not possible in B-DNA. The requirement of helicity constrains the Twist angle to roughly 36° ± 7°.[182] Hence Cooper *et al.* performed an energetic scan of the 10 possible DNA base steps (in the gas phase) with the Twist angle constrained within 0°–60°, using PBE with a van der Waals correction.[195] Despite the absence of solvent and sugar-phosphate backbone, the geometries in which the stacking reached a potential energy minimum as a function of Twist mostly had angles within 5° of the experimentally determined average for each base step. Slightly larger deviations (~15°) were found for a few steps. This was taken as evidence that, within helical constraints, local DNA geometries are mostly determined by the optimal stacking interactions for each distinct base step.

A review by Šponer *et al.* in 2008 presented free energies of base stacking in water derived from CBS(T) potential energies.[120] The results were basically consistent with the experimental findings of Yakovchuk *et al.* and others, varying from −0.3 kcal/mol for the A/C dimer to −1.1 kcal/mol for the G/G dimer. (Lower stacking strengths were found for uracil.) The authors also noted that another quantum method for partitioning the stacking energy into discrete physical terms, Symmetry-Adapted Perturbation Theory (DFT-SAPT), reproduced the earlier finding that dispersion is the dominant attractive force. The review mentioned the discrepancy between, on the one hand, the fact that stacking in the gas phase can now be described very well by high-level *ab initio* methods (and even force-fields, per Ref.[192]), and, on the other hand, that stacking in real DNA varies widely and unpredictably as a function of multiple factors. If there is a set of simple rules relating the intrinsic chemical stability of stacked bases, as revealed by theory, to the actual properties of DNA, as revealed by experiment, it remains elusive.

Morgado *et al.* carried out a large-scale study of over 100 stacked uracil dimers, with subtle geometric variations resembling those found in experimental helical structures.[196] DFT-SAPT confirmed that the dispersion term was large, stabilising and Twist-independent for all structures, while the electrostatic term was Twist-dependent and variably stabilising or destabilising. Interestingly, the authors suggested that the Twist-dependence of the electrostatic term was so great, and so biased towards non-DNA-like Twist values, that the effect of local geometric variations on the stacking of DNA might be more accurately described by completely neglecting the electrostatic interactions, especially as they tend to be heavily screened out by aqueous solvent anyway.

In 2011, Smith *et al*. carried out a decomposition of the stacking energies of A/T and G/C, and their methylated counterparts, using the General Effective Fragment Potential method (EFP2).[15] The contributing energy terms were Coulombic, exchange-repulsion, polarisation, dispersion and charge transfer. The earlier general conclusions were reinforced – dispersion was the largest attractive component for all stacked dimers except non-methylated G/C (for which the Coulombic term

dominated), but the Coulombic term was also crucial for all dimers. Methylation strongly enhanced the stacking of A/T (by increasing both the Coulombic and dispersive contributions) but slightly weakened the stacking of G/C (by decreasing the Coulombic term more than it increased the dispersive term). A comparison with hydrogen-bonded structures repeated the earlier finding by Dąbkowska *et al*.[191] that G/C stacking was stronger than A-T base pairing.

Experimental evidence for the importance of dispersion in aromatic π–π interactions was provided by Lima *et al*. in 2012.[16] Calorimetry of the homodesmotic formation of substituted 1,8-diphenylnaphthalenes revealed a linear relationship between the enthalpies of formation of these systems – whose defining structural feature was closely interacting parallel aromatic rings – and their calculated molecular polarisabilities. The formation reactions were in fact unfavourable for unsubstituted and disubstituted systems, but favourable for the bulkier and hence more polarisable tetrasubstituted systems. The driving force for the formation of the larger compounds was apparently not electrostatic, since it was promoted by both electron-withdrawing and electron-donating substituents. A close correlation between the experimental reaction enthalpies with the calculated dispersion energies of the corresponding benzene derivatives reinforced the conclusion that dispersion was the key force. A similar conclusion was reached by Trnka *et al.* in a theoretical study of parallel-displaced hexahalobenzenes (C_6X_6): DFT-SAPT showed that the increase in interaction energy with increasing halogen atomic number was mostly caused by greater dispersion.[197]

A 2013 review of DNA base stacking by Šponer *et al.* reiterated the point that solvent screening almost entirely cancels out the Coulombic component of stacking, because the arrangement of the water molecules, together with the associated electric field, rewards the electrostatically unfavourable geometries and punishes the electrostatically favourable ones.[198] Among other things, this has the effect of neutralising the strong Twist-dependence of the stacking energy observed in the gas phase. Since DNA *in vitro* is usually studied in aqueous solution, and water is also known to be present in the cellular environment of DNA *in vivo*, this screening effect is a vital caveat to the interpretation of gas-phase calculations that obtain large electrostatic terms.

In summary, the intrinsic *gas-phase* stacking energies of nucleobase dimers in DNA-like geometries lie within the range −5 to −15 kcal/mol. The two main stabilising contributions to this energy are dispersion (contributing roughly −10 to −15 kcal/mol), which is largely independent of the Twist angle, and Coulombic interactions (whose exact contribution is highly Twist-dependent and also theoretical-method-dependent). Therefore, electrostatics determines the optimal Twist in the gas phase. The main repulsive component is the short-range exchange-repulsion (whose magnitude is also highly method-dependent, and largely cancels the Coulombic attraction).

In *solvent*, however, the screening effect largely annihilates the Twist-dependence of electrostatics, and specific base–solvent interactions further complicate the relationship between structure and energy, leading to relatively flat but complicated energy landscapes. Dispersion still makes a large quantitative contribution to the potential energy of stacking in solution, but the major free energetic driving force for the association of nucleobases in water is probably the hydrophobic effect. Free energies of stacking for natural base dimers in solution vary between −0.5 and −1.5 kcal/mol.

5.1.3. Stacking of 5-bromouracil

The idea that the mutagenicity of BrU might be related to its stacking, rather than its tautomeric preferences, is now quite old. It was known by 1962 that the incorporation of BrU raised the melting point of DNA,[199] which could be due to either stronger stacking or stronger base pairing. In 1970 it was shown by spectrophotometry that the tendency of 5-halouracils to associate with adenine in aqueous solution (forming 1:1 complexes that were assumed to be stacked) increased with the van der Waals radius of the 5-substituent, so that BrU stacked with almost twice the equilibrium constant of U.[74] In 1971 it was shown by melting experiments that the enhanced thermal stability of polynucleotides containing BrU was partly sequence-dependent: homopolymers of A•BrU (with one strand made entirely of A and the other entirely of BrU) were more stable than the corresponding copolymers (with alternating A and BrU on each strand).[75] The same effect was observed for A•T polymers, but the melting points were lower. The authors suggested that both the enhanced stability of BrU-containing polynucleotides, and the sequence dependence, may be due to greater stacking caused by the polarisability of BrU, which implies either stronger van der Waals forces or stronger dipole–induced dipole forces.

As mentioned in Chapter 1, in the early 1970s, Sternglanz, Bugg and co-workers developed a stacking model to explain the mutagenicity of BrU.[73, 77, 78] They observed that the crystal structures of 5-halogenated bases contained close contacts between the halogen atom of one base and the aromatic ring of its neighbour, at the expense of reducing the overlap of neighbouring rings compared with non-halogenated bases. Normal B-DNA stacking patterns do not allow such an interaction, because the methyl substituent on thymine points into the groove of the helix rather than stacking directly above the adjacent base. However, they suggested that the incorporation of a 5-halouracil ("XU") resulted in a rearrangement of the stacking geometry, by a slight rotation and translation of the halogenated base, to allow halogen–aromatic short contacts. Whether stacked above a purine or a pyrimidine, the resulting disruption to Watson–Crick stereochemistry would forbid XU from base pairing with adenine, because of a short (2.8 Å) non-bonded contact between O4(XU) and N1(A). A pyrimidine could also not take the place of adenine, as the inter-base separation would be too large to allow hydrogen bonding. However, guanine could base pair with XU by a single hydrogen bond between O4(XU) and N1H(G). The resulting base pair would be mutagenic.

Even if Bugg *et al.*'s model were not accurate in detail, the enhanced stacking of BrU would still arguably constitute indirect support for a model of BrU-G mispairing that does not require tautomerisation. The intrinsic stabilisation of BrU afforded by its stacking would protect *any* BrU-G mispair, whatever its stereochemistry, against removal by repair enzymes, because the energetic penalty for unstacking the base analogue would be greater than for unstacking the natural base, thymine. This argument was advanced by Brown *et al*. in 1986, who found only wobble (i.e. major-tautomeric) structures of the BrU-G mispair in a Z-DNA hexamer, and observed greater overlap of the bases in BrU/G stacks than in the corresponding C/G stacks.[70] The fundamental energetic change caused by the bromine substituent – i.e., the difference between BrU and T which allows the BrU-G mispair to escape excision by the highly sophisticated repair machinery – is here argued to be not the stabilisation of rare tautomers (experimentally hard to verify), but the greater intrinsic stacking strength (which stands on firmer experimental grounds).

A slightly different version of this argument was developed by Petruska *et al.*, also in 1986.[200] Rather than invoke the energy of post-replicational mismatch *repair*, they started from the assumption that

the initial *insertion* of the correct nucleotide occurs with a fidelity that depends on the free energy difference between correct and incorrect base pairs at the polymerase surface. This difference is measured by melting experiments, and can be partitioned into contributions from stacking and hydrogen bonding. Therefore, if BrU takes part in stronger stacking than T, its misinsertion opposite G will be more likely.

The authors then showed, through a correlation between experimental base pair dissociation energies in water and calculated stacking and hydrogen bonding energies in the gas phase, that complete solvation in water reduces the base stacking component of the energy difference to 0.22 of its gas-phase value, and the hydrogen bonding component to essentially zero. This means that stacking is sharply reduced in water, while hydrogen bonding contributes essentially no net stability at all (later confirmed by Yakovchuk *et al*., Ref [18]). Therefore, discriminating between correct and incorrect nucleotide insertion is significantly harder in water than *in vacuo*. Hence it was argued that the high insertion fidelity achieved by DNA polymerases is a function of their ability to remove water from around bases at the active site, thus conserving the vital free energy difference between right and wrong base pairs. However, the water which *is* still present, though in decreased amount, reduces the H-bonding component of fidelity more sharply than it reduces the stacking component (presumably because H-bonding is more electrostatic): and this reinforces the idea that intrinsic stacking mainly determines the insertion of correct nucleotides.

Since base stacking is sequence-dependent, this model of BrU mutagenicity may also explain the sequence specificity of BrU-induced mutations. However, we have not investigated this idea any further.

It is still not known, in any quantitative detail, how much more strongly BrU stacks in DNA compared to the natural bases. BrU was not included by Guckian *et al.* in their dangling-end experiments.[12] Furthermore, it must be remembered that, regardless of whether the responsibility for discriminating against incorrect base pairs is borne by polymerase enzymes at the replication stage, or by repair enzymes at the post-replication stage, or both, the assumption that fidelity is maintained by thermodynamics (rather than stereochemistry) has been challenged.[37, 38, 40]

5.2. Experimental geometries (Py/Py and Py/Pu)

To compare the stacking strengths of 5-bromouracil and thymine, our first approach was to use experimentally-determined geometries of stacked dimers from real DNA structures. The idea was to isolate sequential stacked bases within DNA helices (which we denote X/Y, with the 5' base followed by the 3' base), and determine their gas-phase interaction energies by the supermolecular method. The average interaction energies of BrU and T with each type of stacking partner could then be compared.

The largest and most accessible source of solved DNA structures is the Protein Data Bank (PDB),[201] which contains a large repository of nucleic acid structures as well as proteins. Most structures are determined by X-ray diffraction, while a minority are determined by proton NMR. Each entry in the PDB is identified by a four-character alphanumeric code.

We searched the PDB for every nucleic acid entry featuring the string "BrU". In total, we located 28 DNA structures containing BrU nucleotides. From these structures, we isolated 47 distinct stacked

dimers containing BrU stacking with either A, C, G or T. For comparison, we also isolated 51 dimers containing T stacking with A, C, G or T, from 17 separate PDB files (four of which were also among the BrU-containing entries). This gave 98 structures in total. All the PDB codes used to generate structures in this study are listed in Appendix B.

The stacked dimers were created by isolating two adjacent nucleotides within a strand (i.e. deleting all the other nucleotides, as well as any water molecules, metal ions and polypeptides), then converting the nucleotides into nucleobases, by deleting all the sugar-phosphate backbone atoms. Since the PDB coordinates do not include hydrogen atoms, we manually saturated each base by placing H atoms in positions corresponding to the canonical tautomers.

We then geometry-optimised the positions of all the H atoms, while keeping the other atoms frozen, using the "ModRedundant" keyword in Gaussian 09.[136] These partial optimisations were performed on the stacked nucleobase dimer systems (rather than separately on each monomer), at the M06-2X/6-31+G(d) level of theory.

We then calculated counterpoise-corrected single-point interaction energies of each of the resulting dimer geometries, at a range of levels of theory. The counterpoise-corrected interaction energy (stacking energy), E_{int}, was defined as

$$E_{\mathrm{int}} = E_{\mathrm{X/Y}}^{\{\mathrm{X/Y}\}}(\mathrm{X/Y}) - E_{\mathrm{X}}^{\{\mathrm{X/Y}\}}(\mathrm{X/Y}) - E_{\mathrm{Y}}^{\{\mathrm{X/Y}\}}(\mathrm{X/Y}) \qquad (5.1)$$

where the terms on the right-hand side are the energies of the dimer, the 5' base and the 3' base, respectively, all calculated at the dimer geometry, in the superposed basis set of the dimer ("dimer-centred basis set"). The energies of each isolated base were calculated in the presence of ghost orbitals associated with the other base. This "vertical" counterpoise method takes into account the basis set superposition error, but not the deformation energies, which we decided to neglect due to the use of experimental base geometries.

It is important to note that stacked dimers of sequence 5'-X/Y-3' and 5'-Y/X-3' are not geometrically equivalent if X≠Y, due to the structure of the helix. If the geometry of a dimer is expressed as the displacement of Y relative to X, then the Slide and Twist (the translation of Y along the *y*-axis and its rotation about the *z*-axis) proceed in opposite senses in the 5'-X/Y-3' and 5'-Y/X-3' dimers. Due to the lack of certain symmetry elements in the natural bases, these two different senses result in different displacements of the functional groups of X and Y, relative to the case with Slide = 0 Å and Twist = 0°. This in turn causes different stacking interactions between X and Y. To take this into account, we will henceforth subdivide our 98 stacked dimers into 15 distinct sequences: BrU/A, A/BrU, BrU/C, C/BrU, BrU/G, G/BrU, BrU/T, T/BrU, T/A, A/T, T/C, C/T, T/G, G/T and T/T. The number of data points (experimental structures) for each sequence is 8, 8, 4, 4, 7, 8, 5, 3, 5, 9, 8, 5, 7, 7 and 10, respectively.

5.2.1. MP2

The first level of theory chosen for calculating the interaction energies was MP2/6-31+G(d). These calculations were performed using Gaussian 09. The results are shown in Table 5.2.

Table 5.2 Mean MP2/6-31+G(d) interaction energies ΔE^{mean} (kcal/mol) of different types of experimental stacked dimer

Type X/Y	ΔE^{mean}	Type Y/X	ΔE^{mean}	All X + Y[a]	ΔE^{mean}
BrU/A	−5.44	A/BrU	−5.62	BrU + A	−5.53
T/A	−4.15	A/T	−5.21	T + A	−4.83
BrU/C	−3.48	C/BrU	−6.38	BrU + C	−4.93
T/C	−2.66	C/T	−3.92	T + C	−3.14
BrU/G	−4.87	G/BrU	−4.23	BrU + G	−4.53
T/G	−3.98	G/T	−3.54	T + G	−3.76
BrU/T	−1.60	T/BrU	−3.53	BrU + T	−2.32
T/T	−1.80	T/T	−1.80	T + T	−1.80

X = T or BrU; Y = A, C, G or T.
[a] With X/Y and Y/X grouped together.

To allow comparison with both T/BrU and BrU/T, the average T/T energies are shown twice.

The BrU-containing dimers are, on average, more strongly bound than the corresponding T-containing dimers, except when comparing the mean of 5'-T/T-3' with that of 5'-BrU/T-3'. The largest difference between means is for 5'-C/X-3', where the mean interaction energy for X = BrU is −6.38 kcal/mol (calculated from four structures), while the mean for X = T is −3.92 kcal/mol (calculated from five structures), i.e. the difference is 2.46 kcal/mol in favour of BrU.

The final column shows the mean stacking energies of X/Y and Y/X grouped together, i.e. for each combination (X + Y), the difference between having X on the 5' side or the 3' side is ignored. In this case, the mean BrU + Y stacking strength is greater than the mean T + Y stacking strength for every Y, with the biggest difference being BrU + C (−4.93 kcal/mol) vs. T + C (−3.14 kcal/mol).

The least favourable interaction energy for any individual structure was +1.03 kcal/mol (for 3OH9-T/T, this being one of two dimers with positive stacking energies), while the most favourable was −8.54 kcal/mol (for 1DCR-BrU/G). The effect of the BSSE on these energies was significant, being of the same order of magnitude as the final interaction energies. The BSSE corrections (not shown in the table) ranged from 2.50 kcal/mol (for 1AIO-T/G) to 11.48 kcal/mol (for 1A35-A/BrU).

5.2.2. df-MP2 combined with df-LCCSD(T)

MP2 is not only highly susceptible to BSSE, it also systematically over-binds aromatic stacked dimers, as revealed by higher-level methods such as coupled cluster.[120, 192] We therefore decided to repeat these single-point calculations with coupled cluster. The "gold standard" of quantum chemistry is CCSD(T)/CBS. For calculating comparative stacking energies, we decided that the aug-cc-pVQZ basis set (henceforth "aQZ") should be sufficiently large. For systems of this size (~30 atoms), direct calculation of CCSD(T)/aQZ was too computationally demanding for the hardware available to us. However, the combination of coupled cluster with a large basis set can be indirectly estimated by a method called focal-point analysis.[202] The coupled cluster correction – the difference between CCSD(T) and MP2 energies for a given basis set – is known to converge faster, with respect to basis set size, than the absolute coupled cluster energies. The correction can therefore be calculated

in a small basis set, then added to a large-basis MP2 energy to obtain the large-basis CCSD(T) energy:

$$E_{\mathrm{CCSD(T)}}^{\mathrm{large}} \approx E_{\mathrm{MP2}}^{\mathrm{large}} + \left[E_{\mathrm{CCSD(T)}}^{\mathrm{small}} - E_{\mathrm{MP2}}^{\mathrm{small}}\right] \tag{5.2}$$

where the final two terms give the coupled cluster correction. Our choice of small basis set was aDZ. However, for our stacked dimers, CCSD(T)/aDZ still proved intractable, due to disk space requirements. We therefore decided to use density fitting (Chapter 2, Section 2.8) to reduce the computational demands. In the case of coupled cluster, density fitting only affords significant computational savings when applied in conjunction with the local orbital approximation[203] (Chapter 2, Section 2.9). We therefore used the df-LCCSD(T)/aDZ level of theory. For MP2, we used density fitting, but not the local approximation. Equation 5.2 therefore becomes:

$$E_{\mathrm{df-LCCSD(T)}}^{\mathrm{aQZ}} \approx E_{\mathrm{df-MP2}}^{\mathrm{aQZ}} + \left[E_{\mathrm{df-LCCSD(T)}}^{\mathrm{aDZ}} - E_{\mathrm{df-MP2}}^{\mathrm{aDZ}}\right] \tag{5.3}$$

All single-point calculations were carried out using Molpro 2010.1.[204] At each explicitly calculated level of theory (df-MP2/aDZ, df-MP2/aQZ and df-LCCSD(T)/aDZ), the stacking energies of each dimer were evaluated using Equation 5.1. That is, at each level of theory, three quantities were calculated for each dimer – the energies of X/Y, X and Y – all using the dimer-centred basis set. The vertical counterpoise correction was accomplished in this case by using the "Dummy" keyword in Molpro to specify the ghost atoms. Finally, the CP-corrected stacking energies for df-LCCSD(T)/aQZ were estimated using Equation 5.3.

The resulting interaction energies at all four levels of theory are shown in Table 5.3.

Table 5.3 Mean df-MP2/aug-cc-pVDZ, df-MP2/aug-cc-pVQZ, df-LCCSD(T)/aug-cc-pVDZ and estimated df-LCCSD(T)/aug-cc-pVQZ interaction energies (kcal/mol) for each type of experimental dimer

Type	ΔE^{mean}(df-MP2)		ΔE^{mean}(df-LCCSD(T))	
	aug-cc-pVDZ	aug-cc-pVQZ	aug-cc-pVDZ	aug-cc-pVQZ
Type X/Y				
BrU/A	−7.05	−7.54	−6.19	−6.68
T/A	−5.85	−6.45	−4.85	−5.44
BrU/C	−5.03	−5.42	−4.09	−4.48
T/C	−4.71	−5.27	−3.73	−4.30
BrU/G	−6.29	−6.78	−5.46	−5.96
T/G	−5.41	−5.78	−4.97	−5.34
BrU/T	−4.01	−4.84	−3.10	−3.93
T/T	−4.04	−4.68	−3.18	−3.81
Type Y/X				
A/BrU	−8.19	−9.10	−7.02	−7.93
A/T	−7.74	−8.46	−7.38	−8.10
C/BrU	−8.15	−8.88	−7.15	−7.87
C/T	−6.10	−6.72	−5.41	−6.04
G/BrU	−6.67	−7.39	−5.54	−6.26
G/T	−5.92	−6.70	−4.77	−5.55
T/BrU	−5.49	−6.13	−4.74	−5.38
T/T	−4.04	−4.68	−3.18	−3.81
All X + Y[a]				
BrU + A	−7.62	−8.32	−6.60	−7.31
T + A	−7.06	−7.74	−6.48	−7.15
BrU + C	−6.59	−7.15	−5.62	−6.18
T + C	−5.24	−5.83	−4.38	−4.97
BrU + G	−6.49	−7.11	−5.50	−6.12
T + G	−5.66	−6.24	−4.87	−5.44
BrU + T	−4.56	−5.32	−3.72	−4.48
T + T	−4.04	−4.68	−3.18	−3.81

X = T or BrU; Y = A, C, G or T.

[a] With X/Y and Y/X grouped together.

With df-MP2/aDZ, the mean BrU stacking is stronger than the mean T stacking for all comparisons except 5'-X/T-3', for which the energy difference is < 0.05 kcal/mol in favour of T. The largest difference in favour of BrU is 2.05 kcal/mol, for 5'-C/X-3'. The individual interaction energies ranged from −1.78 kcal/mol (3OH9-T/T) to −10.32 kcal/mol (1DCR-BrU/G). These were also the least and most strongly bound dimers, respectively, with MP2/6-31+G(d).

By comparison with Table 5.2, it can be seen that the df-MP2/aDZ interaction energies are already larger than the MP2/6-31+G(d) energies, presumably due to the greater size of this basis set (even

though they are both double-ζ). We attempted MP2/aDZ calculations (without density fitting) for comparison, but they proved impracticable, due to disk space requirements.

With df-MP2/aQZ, the mean BrU stacking is stronger than the mean T stacking for all comparisons. The largest mean difference in favour of BrU is 2.16 kcal/mol, for 5'-C/X-3'. The mean interaction energies are larger than with the aDZ basis set. The individual interaction energies ranged from −2.36 kcal/mol (2L5K-G/T) to −11.13 kcal/mol (1DCR-BrU/G).

With df-LCCSD(T)/aDZ, the mean BrU stacking is stronger than the mean T stacking for all comparisons except 5'-A/X-3', where the difference in favour of T is 0.36 kcal/mol, and 5'-X/T-3', for which the mean difference is < 0.1 kcal/mol. The largest mean difference in favour of BrU is 1.74 kcal/mol, for 5'-C/X-3'. The individual interaction energies ranged from +0.33 kcal/mol (2W7N-BrU/T) to −9.79 kcal/mol (2L5K-A/T).

With df-LCCSD(T)/aQZ – the highest level of theory – the mean BrU stacking is stronger than the mean T stacking for all comparisons except 5'-A/X-3', where the difference in favour of T is 0.17 kcal/mol. The largest mean difference in favour of BrU is 1.83 kcal/mol, for 5'-C/X-3'. The individual interaction energies ranged from −0.93 kcal/mol (2W7N-BrU/T) to −10.73 kcal/mol (2L5K-A/T).

With all four methods, when X/Y and Y/X are grouped together, the only partner base (Y) for which the mean stacking energy of BrU exceeds that of T *by more than 1 kcal/mol* is cytosine.

It must be noted that there is an inconsistency between the two Hamiltonians employed: df-LCCSD(T) uses localised orbitals, while df-MP2 does not. This may call into question the accuracy of our coupled cluster correction, the addition of which to the df-MP2/aQZ energies yields the df-LCCSD(T)/aQZ energies. It was not possible to avoid this mismatch by using df-CCSD(T) or df-LMP2. The former yields negligible computational savings compared to full coupled cluster (as noted above),[203] and therefore requires more computer memory than we had at our disposal. The latter, which should in principle be computationally cheaper than df-MP2, also proved impracticable on our hardware, for reasons apparently related to memory.

In a review of theoretical studies of stacking, Šponer *et al.*[120] noted that the coupled cluster correction, ΔCC, is usually large and positive (repulsive). In our own calculations, ΔCC is on average positive for each type of dimer (compare the second and fourth columns in Table 5.3), but it is negative for eight specific dimers – in one case by over 1 kcal/mol (the A/T dimer from 2L5K, for which ΔCC = −1.4 kcal/mol). However, Sponer *et al.*[192] had earlier calculated CCSD(T) corrections in the range of −0.1 to +2.5 kcal/mol for DNA base pair steps, i.e. ΔCC can occasionally be weakly attractive.

To estimate the density fitting error in our results, we calculated the Hartree–Fock total energies of two dimers (1D9R-BrU/G and 1A35-A/T) with and without density fitting. The differences were small (~0.2 kcal/mol) in the aDZ basis and negligible (<0.05 kcal/mol) in the aTZ and aQZ bases.

5.2.3. RI-mPW2PLYP-D

The double hybrid functional RI-mPW2PLYP-D consists of the mPW2PLYP[114] functional (modified Perdew–Wang exchange with 55% contribution from Hartree–Fock, and Lee–Yang–Parr correlation

with 25% contribution from RI-PT2), supplemented with the 2006 version of Grimme's D2 empirical dispersion correction.[117]

The RI-mPW2PLYP-D calculations were carried out using Orca 2.8.[205] The keyword "vdw" was used to invoke Grimme's D2-type dispersion correction. The counterpoise correction was applied by specifying ghost atoms in the geometry inputs, and the interaction energy was defined as above.

The interaction energies are shown in Table 5.4.

Table 5.4 Mean RI-mPW2PLYP-D/aug-cc-pVTZ interaction energies ΔE^{mean} (kcal/mol) of different types of experimental stacked dimer

Type X/Y	ΔE^{mean}	Type Y/X	ΔE^{mean}	All X + Y[a]	ΔE^{mean}
BrU/A	−6.40	A/BrU	−6.56	BrU + A	−6.48
T/A	−5.37	A/T	−6.27	T + A	−5.95
BrU/C	−4.15	C/BrU	−7.63	BrU + C	−5.89
T/C	−3.88	C/T	−5.65	T + C	−4.56
BrU/G	−5.57	G/BrU	−4.88	BrU + G	−5.20
T/G	−4.64	G/T	−4.72	T + G	−4.68
BrU/T	−2.97	T/BrU	−4.34	BrU + T	−3.48
T/T	−3.16	T/T	−3.16	T + T	−3.16

X = T or BrU; Y = A, C, G or T.
[a] With X/Y and Y/X grouped together.

The mean BrU stacking is stronger than the mean T stacking for all comparisons except 5'-X/T-3', for which the difference is 0.19 kcal/mol in favour of T. However, the differences are mostly small. The largest mean difference in favour of BrU is 1.98 kcal/mol, for 5'-C/X-3'. The individual interaction energies ranged from −0.28 kcal/mol (2W7N-BrU/T) to −9.81 kcal/mol (1DCR-BrU/G). When X/Y and Y/X are grouped together (final column), the only partner base for which the mean stacking energy of BrU exceeds that of T *by more than 1 kcal/mol* is cytosine.

5.3. Twist angle scanning and full geometry optimisations (Py/Py only)

The above calculations on experimentally-derived structures suggest that, in general, stacking potential energies are similar for BrU and T. However, the 5'-C/BrU-3' stacking interaction is around 2 kcal/mol stronger than the 5'-C/T-3' interaction – a potentially significant difference. At each level of theory, this pair of dimers is the one with the greatest stacking energy difference in favour of BrU. Interestingly, however, the mean stacking energies of 5'-BrU/C-3' and 5'-T/C-3' – the same bases, but in reverse order – are essentially identical at all levels of theory. To complement and further interpret these results, we decided to move beyond fixed stacking geometries, and explore in more detail the conformational spaces of BrU, T and U stacking with the other pyrimidines.

The basis for our methodology was an earlier computational study by Hunter and van Mourik, which explored the potential energy surfaces of the uracil dimer (U/U) and the thymine dimer (T/T).[206] In

that study, the two bases were placed one exactly atop the other, at a vertical separation of 3.4 Å, then one base was rotated around its geometric centre while the other was kept fixed, by varying an angle, τ, analogous to the Helical Twist between base pairs. The counterpoise-corrected M06-2X/6-31+G(d) interaction energies were calculated at regular increments of this angle to generate potential energy curves for rotation, and the structures corresponding to minima on these curves were then fully geometry-optimised at the same level of theory. Transition states between the resulting minima were also located.

For each dimer, the authors considered two distinct orientations: face-to-back (FTB, one base exactly aligned with the other) and face-to-face (FTF, with one base flipped by 180° around the C2–C5 axis, so that its N1 atom lies above the N3 atom of the other base and *vice versa*). Due to the lack of certain symmetry elements in the natural bases, these two orientations are geometrically distinct, and give rise to different stacking interactions and rotational profiles. In B-DNA, intra-strand stacking is always face-to-back (due to the requirements of Watson–Crick base pairing), but face-to-face stacking has been observed in inter-strand interactions.[207, 208]

For U/U, Hunter and van Mourik calculated the FTB and FTF rotational energy profiles in the gas phase, and with two different PCM solvents – water and 1,4-dioxane – while for T/T, only the gas phase and water were considered. For U/U, the global energy minimum in the gas phase was an FTF structure with $\tau = 48°$ and CP-corrected interaction energy $\Delta E^{CP} = -9.78$ kcal/mol. For T/T, the global energy minimum in the gas phase was an FTF structure with $\tau = 41°$ and CP-corrected interaction energy $\Delta E^{CP} = -11.04$ kcal/mol. For U/U, the most strongly bound FTB dimer in the gas phase had $\tau = 180°$ and $\Delta E^{CP} = -9.34$ kcal/mol, while for T/T the equivalent values were $\tau = 70°$ and 290° and $\Delta E^{CP} = -10.30$ kcal/mol. The effect of both solvents was to flatten the gas-phase potential energy profiles, by reducing the magnitudes of the interaction energies and lowering the barrier heights between adjacent minima (as defined by the transition state energies). The U/U FTF stacking energy of −9.78 kcal/mol was in excellent agreement with CCSD(T)/CBS estimates for this structure, which range from −9.74 to −9.83 kcal/mol.[202, 206]

However, it was clear that the potential energy surfaces of isolated, stacked base dimers were quite different from those of stacked dinucleotides in DNA (i.e. connected by a sugar-phosphate backbone): when Hunter and van Mourik fully optimised their dimers, no FTB energy minima were found for either U/U or T/T that had a value of τ within 20° of the canonical B-DNA value of the Helical Twist (36°).

We decided to repeat their methodology for the following dimers: U/BrU, BrU/BrU, C/BrU, T/BrU, C/C, C/T, C/U and T/U. For all eight systems, the gas phase and a water (PCM) environment were considered; 1,4-dioxane (PCM) was included for U/BrU only. The procedure will now be described in detail.

The stacked dimers, here denoted A/B, were created by placing the two base molecules in a parallel configuration, using Molden.[209] Each base was in its monomer-optimised geometry (from the gas phase or PCM solvent, as appropriate). Face-to-back and face-to-face arrangements of the bases were both constructed. Dummy atoms (Z) were placed in the geometric centres of each ring, as calculated from the C2, C4 and C6 atoms. The vertical separation between the bases was chosen as 3.4 Å, as this is a typical value of the Rise base step parameter in B-form DNA, according to El Hassan and

Calladine[183] (note that Olson *et al.*, whose data are shown in Table 5.1, give a slightly smaller average Rise of 3.32 Å).[182]

Rotational energy profiles were then calculated by rotating one base around the geometric center of its ring, while keeping the other fixed, by varying the O2(A)–Z(A)–Z(B)–O2(B) torsion angle (τ) between 0–355° with a step size of 5°. This was accomplished using the "Scan" keyword in Gaussian 09. All other geometrical parameters were held constant.

Interaction energies, ΔE^{CP}, were calculated at each step at the M06-2X/6-31+G(d) level of theory. These energies were obtained as follows:

$$\Delta E^{\mathrm{CP}} = E_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) \tag{5.4}$$

The solution was modelled by the polarisable continuum model (PCM), using the integral equation formalism variant (IEFPCM), via the "SCRF" keyword in Gaussian 09. The cavity containing the solute was constructed by placing spheres around each solute atom (including all hydrogens) using universal force field (UFF) radii.

The unrelaxed geometries giving rise to minima in the potential energy scans were subsequently fully optimised at the same level of theory, M06-2X/6-31+G(d). Unlike with the experimental dimers and the scans, the interaction energies of the optimised dimers included the deformation energies as well as the "vertical" counterpoise corrections. The final interaction energies were thus

$$\begin{aligned}\Delta E^{\mathrm{CP}} = {} & E_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) + E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B}) + E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B}) \\ & - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B})\end{aligned} \tag{5.5}$$

where the additional four terms (compared to Equation 5.4) account for the deformation energies. Due to the use of monomer-optimised geometries for twist-angle scanning, the deformation energies had been zero during the scans.

Transition states were then located by the Synchronous Transit-Guided Quasi-Newton (STQN) method,[210] using the "Opt = QST3" keyword in Gaussian 09. This method requires specifying three input geometries: two minima (for which we used the optimised structures), and one guess geometry for the transition state connecting them (for which we used the geometries that yielded energy maxima in the rigid scans). Sometimes, the QST3 method failed to locate a stationary point after exceeding the allowed number of optimisation steps, or encountering some other error – in these cases, the final geometry from the QST3 search was used as the starting geometry for a conventional transition state search using the "Opt = TS" keyword, together with the "NoEigenTest" option to suppress testing of the curvature. One or other method always successfully located a transition state. All optimisations and transition state searches were carried out using Gaussian's "Tight" convergence criteria. All calculations employed Gaussian's ''Ultrafine'' integration grid (containing 99 radial shells and 590 angular points per shell). Harmonic vibrational frequencies were computed at the same levels of theory, to verify the nature of the stationary points (minima or transition states), and to compute zero-point energy (ZPE) and thermal corrections to the interaction energies.

In the gas phase, the counterpoise correction was applied throughout the geometry optimisations and transition state searches as well as during the rigid scans. In the PCM calculations, the counterpoise

correction was applied only to the rigid scans and final optimised structures, in separate calculations without the "SCRF" keyword, as it is not possible to implement counterpoise in a PCM calculation.

For each optimised structure, we then computed four geometrical parameters, using an in-house Fortran 90 program written by Dr. Tanja van Mourik, which was previously described in Ref. [206] Each of these parameters is analogous to one of the base step parameters defined in Section 5.1, but they are not formally equivalent, as our parameters are defined for stacked dimers rather than complete base steps.

The first parameter is the dimer twist angle, analogous to Helical Twist. It is simply the final optimised value of the O2(A)–Z(A)–Z(B)–O2(B) torsion angle, τ, which was varied during the scans.

The second parameter is the vertical distance between the two stacked bases, analogous to Rise. The vertical distance was calculated as the scalar product of the vector connecting the midpoints of the base rings ($\overrightarrow{m_1 m_2}$; calculated as the average of the atomic coordinates of the C2, C4 and C6 atoms) and the plane vector of one of the bases (where a plane is defined by the C2, C4 and C6 atoms). This corresponds to the component of $\overrightarrow{m_1 m_2}$ that is perpendicular to the base plane. As the two bases are not completely parallel in the optimised structures, this scalar product depends on which base's plane vector is used. The vertical distance presented herein is therefore computed as the average of the two values obtained when using the first or the second base's plane vector.

The third parameter is the horizontal displacement from the position where one base is exactly on top of the other, analogous to Slide. The horizontal distance is calculated as the component of the vector connecting the two midpoints ($\overrightarrow{m_1 m_2}$) that is in the plane of one base. Like for the vertical separation, the horizontal separation is calculated as the average of the two values obtained when using the first or the second base's plane vector.

The fourth parameter is the angle between the two planes of the optimised stacked dimers, analogous to Tilt. The angle between the planes of the two bases is simply the angle between their plane vectors.

The fixed values of these parameters during the twist scans had been as follows: vertical separation = 3.4 Å, horizontal separation = 0 Å, angle between planes = 0°.

To take into account thermal effects, we calculated the enthalpic and free energetic corrections to the potential energies, at T = 298.15 K, using Gaussian 09.

$H_{\mathrm{vib}}(T)$ was calculated for each stacked dimer (A/B), as well as for the isolated monomers (A and B), as outlined in Chapter 2, Section 2.13. For each dimer and monomer (generically "X"), the enthalpy, H, was then calculated as the sum of the SCF potential energy and the vibrational enthalpic correction:

$$H_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) = E_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) + H_{\mathrm{vib,X}}(T) \qquad \mathrm{X} = \mathrm{A}, \mathrm{B}, \text{or } \mathrm{A/B} \tag{5.6}$$

The counterpoise-corrected enthalpies of stacking, ΔH^{CP}, are then given by

$$\Delta H^{\mathrm{CP}} = H_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - \mathrm{BSSE} - H_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - H_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B}) \tag{5.7}$$

where the BSSE is calculated from the SCF potential energies as before:

$$\mathrm{BSSE} = E_{\mathrm{A}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) + E_{\mathrm{B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - E_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A/B}) - E_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{A/B}) \quad (5.8)$$

Likewise, the vibrational entropic corrections, $S_{\mathrm{vib}}(T)$, were obtained using the frequencies. As with the enthalpy, $S_{\mathrm{vib}}(T)$ was calculated for each stacked dimer (A/B) and also for the isolated monomers (A and B).

The Gibbs free energies, G, of each system are then given by:

$$G_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) = H_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) - TS_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) \quad (5.9)$$

where $S_{\mathrm{X}}^{\{\mathrm{X}\}}(\mathrm{X}) \equiv S_{\mathrm{vib}}(T)$ for an isolated monomer or a stacked dimer. The counterpoise-corrected stacking free energies, ΔG^{CP}, are then given by

$$\Delta G^{\mathrm{CP}} = G_{\mathrm{A/B}}^{\{\mathrm{A/B}\}}(\mathrm{A/B}) - \mathrm{BSSE} - G_{\mathrm{A}}^{\{\mathrm{A}\}}(\mathrm{A}) - G_{\mathrm{B}}^{\{\mathrm{B}\}}(\mathrm{B}) \quad (5.10)$$

We also calculated the isotropic molecular polarisabilities, α, of the individual bases, at the M06-2X/6-31+G(d) level of theory, both in the gas phase and in water PCM. These data are given in Table 5.5. Recall that there is an experimentally observed relationship between polarisability and stacking strength.[12, 16]

	5-Bromouracil	Thymine	Uracil	Cytosine
α (gas) / bohr^3	83.2	76.3	64.2	71.2
α (water) / bohr^3	110.4	100.4	84.5	94.3

Table 5.5. Isotropic polarisabilities, α, of BrU, T, U and C, calculated in the gas phase and water (PCM) at the M06-2X/6-31+G(d) level of theory.

The main results section contains a large amount of data. First, the geometric parameters, interaction energies and interaction enthalpies of the optimised minima and transition states will be tabulated for each dimer. For ease of reading, the enthalpies will only be given for the energy minima, not the transition states. The free energies will be introduced later, in the Discussion (Section 5.4). For U/BrU, but not for the other dimers, the interaction energy profiles will also be shown here.

5.3.1. U/BrU

Fig. 5.2 shows the M06-2X/6-31+G(d) gas-phase interaction energy of the face-to-back U/BrU stacked dimer as a function of the τ_{twist} angle. The BSSE was found to be fairly large, varying between 3.05 kcal/mol (when $\tau_{\mathrm{twist}} = 320°$) and 3.67 kcal/mol (when $\tau_{\mathrm{twist}} = 95°$). Energy minima are located at $\tau_{\mathrm{twist}} = 85°$, 170°, 205° and 290°. Energy maxima are located at $\tau_{\mathrm{twist}} = 5°$, 120°, 185° and 240°. The interaction energies around the highest maximum, at 5°, are positive (repulsive), presumably due to the use of a rigid scan. Note that all the fully optimised minima and transition states, including the one near 5° (with $\tau_{\mathrm{twist}} = 9°$), have negative interaction energies (Table 5.6).

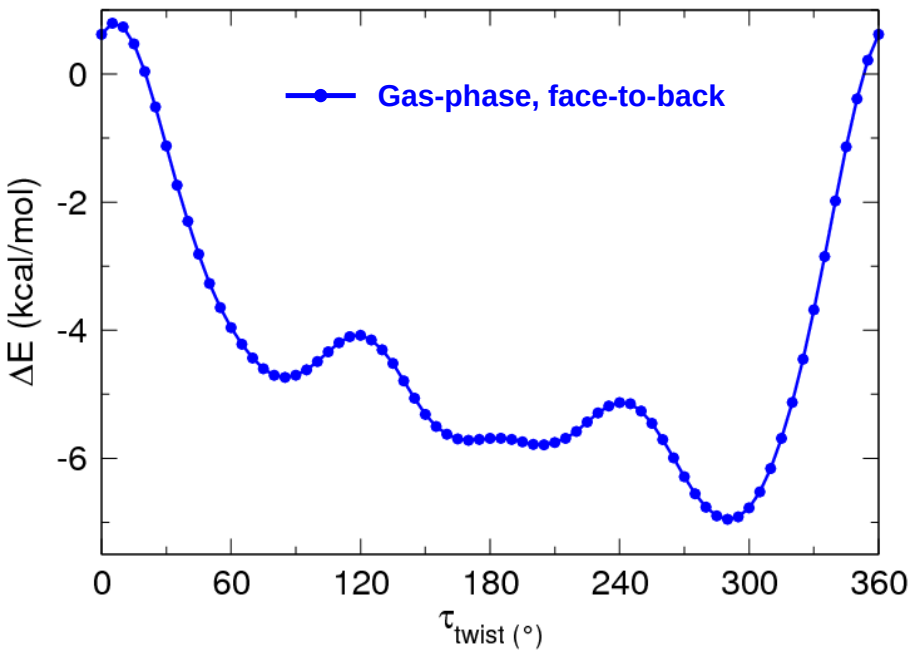


Figure 5.2: Interaction energies as a function of the twist angle for the gas-phase face-to-back U/BrU stacked dimer calculated at the M06-2X/6-31+G(d) level of theory.

Fig. 5.3 shows the interaction energies of the face-to-back U/BrU stacked dimer in PCM solvent (water and 1,4-dioxane) as a function of the τ_{twist} angle. The BSSE must be calculated without taking into account the influence of the solvent. In both solvents there are four minima, located at τ_{twist} = 90°, 155°, 210° and 295° in water and 90°, 165°, 205° and 290° in 1,4-dioxane. The maxima are at 0°, 125°, 180° and 245° in water and 5°, 120°, 185° and 245° in 1,4-dioxane.

The water curve is much flatter than the gas and 1,4-dioxane curves, with only 2.3 kcal/mol separating the highest maximum (at 0°) from the deepest minimum (at 295°). As mentioned in Section 5.1.2., this reduction of the twist-dependence of stacking in polar solvent has been observed by other researchers, and is due to solvent screening, which contributes an energy term that is stabilising for geometries with unfavourable electrostatic interactions and *vice versa*, and thus largely nullifies the electrostatic contribution to stacking.[198]

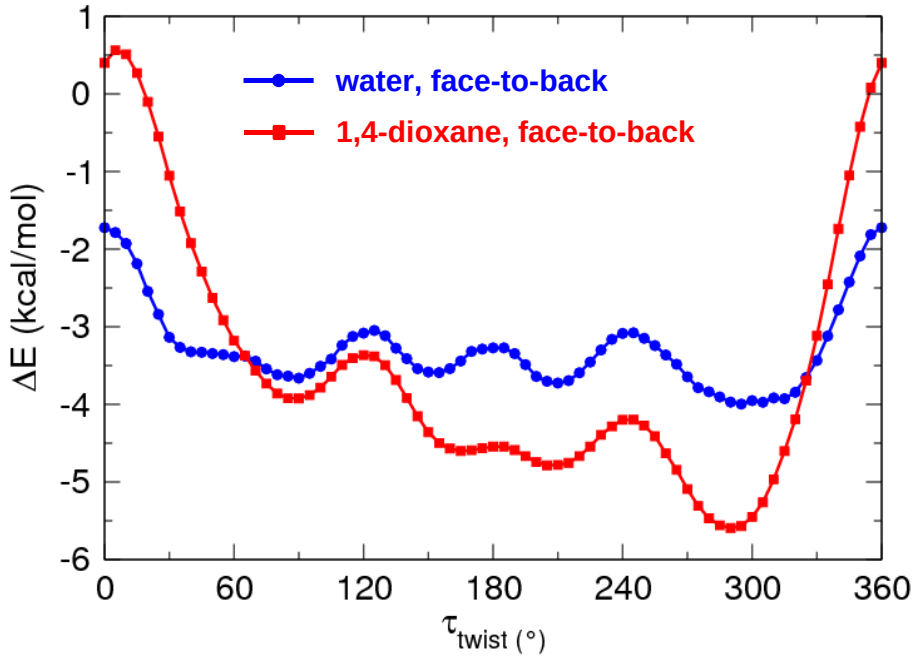


Figure 5.3: Interaction energies as a function of the twist angle for the face-to-back U/BrU stacked dimer in PCM solvent (water and 1,4-dioxane) calculated at the M06-2X/6-31+G(d) level of theory.

Full geometry optimisation of the four gas-phase minima yielded only three distinct structures, since optimisation from τ_{twist} = 165° and τ_{twist} = 205° yielded identical structures. All three optimised minima were confirmed as true minima by the absence of negative frequencies of the calculated harmonic vibrational modes. Table 5.6 shows the interaction energies and geometrical parameters of all face-to-back minima (and transition states) in the gas phase. All interaction energies are negative, including those of the transition states. The most stable structure has a twist angle of 287°, and its CP-corrected, ZPE-corrected M06-2X/6-31+G(d) interaction energy is −9.8 kcal/mol. However, it is almost isoenergetic with the minimum at τ_{twist} = 184°, which has an interaction energy of −9.5 kcal/mol. The third minimum is less stable (ΔE_0^{CP} = −7.2 kcal/mol). None of these structures are perfectly planar or parallel, as can be seen from the non-zero values of the horizontal separation of the two bases and the angles between their planes, respectively. All three minima also exhibit a decrease in vertical separation, to around 3.1 Å, compared with the fixed value of 3.4 Å in the rotational scans.

Table 5.6 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back U/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
5	9	18	2.94	3.72	1	−5.7	−5.3	
85	68	20	3.12	0.91	0	−7.9	−7.2	−3.4
120	189	4	3.15	1.92	1	−8.3	−7.7	
120[a]	183	2	3.00	2.37	0	−8.9	−8.2	−4.0
165 & 205	184	9	3.09	1.51	0	−10.1	−9.5	−3.9
240	226	6	3.22	0.69	1	−7.5	−7.0	
290	287	12	3.06	0.89	0	−10.6	−9.8	−4.7

[a] This minimum was obtained by "following" the imaginary frequency of the transition state listed above it (see text)

Transition states (TSs) were successfully located for three of the four maxima in the gas-phase energy profile. However, the TS obtained from τ_{twist} = 120° (optimised twist: 189°) has an interaction energy of −7.7 kcal/mol, more negative than the minimum obtained with τ_{twist} = 68° (−7.2 kcal/mol), suggesting that it does not connect that minimum with any other. By "following the imaginary frequency" – performing geometry optimisations on structures created from this TS by adding or subtracting the magnitudes of the displacement of each atom in the imaginary vibrational mode – two minimum-energy structures were obtained: one was identical to the minimum with τ_{twist} = 184°, while the other was similar to that one (having τ_{twist} = 183°) but with a larger horizontal separation (2.37 Å as opposed to 1.51 Å) and a stabilization energy of −8.2 kcal/mol. This structure thus represents a fourth minimum on the U/BrU gas phase face-to-back stacking energy surface. It is presumably linked to the minimum with τ_{twist} = 184° *via* the TS with τ_{twist} = 189°, which has an intermediate horizontal separation of 1.92 Å.

Four energy minima and four transition states were obtained in the case where water was used as a

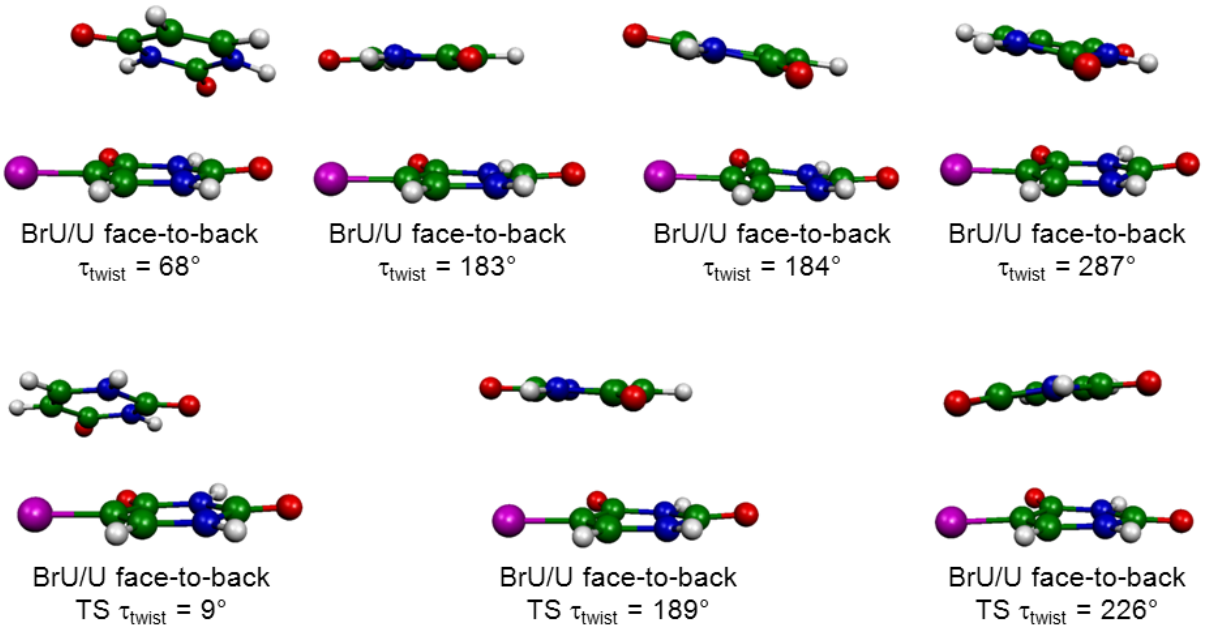


Figure 5.4: Structures of the stacked face-to-back BrU/U minima (top) and transition states (bottom) optimised in the gas phase at the M06-2X/6-31+G(d) level of theory.

PCM solvent. Their geometrical parameters and interaction energies are listed in Table 5.7. The most stable minimum has an optimised twist of 293° and an interaction energy of −5.2 kcal/mol, *i.e.* it is geometrically similar to the gas-phase global minimum but less tightly bound. Note that the TS obtained from τ_{twist} = 245° has an optimised twist angle of 313°, which does not lie between those of the minima it apparently connects (242° and 293°).

Table 5.7 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back U/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	331	10	3.10	2.24	1	−4.3	−4.0	
90	67	6	3.15	1.41	0	−4.4	−4.0	−3.4
125	122	1	3.10	1.18	1	−4.3	−4.0	
155	130	2	3.16	1.19	0	−5.1	−4.7	−4.0
180	182	1	3.24	0.23	1	−3.5	−3.3	
210	242	3	3.16	1.13	0	−4.6	−4.1	−3.4
245	313	5	3.17	1.87	1	−3.8	−3.6	
295	293	6	3.09	1.12	0	−5.5	−5.2	−4.5

The situation with the optimised face-to-back structures in 1,4-dioxane, as shown in Table 5.8, is more complicated. Only three distinct energy minima were found, because optimisation from τ_{twist} = 205° yielded the same structure as starting from τ_{twist} = 290°, with an optimised twist of 288°. Optimisation from the minimum with τ_{twist} = 90° originally yielded a structure with one imaginary frequency, but "following" this frequency with one set of imaginary atomic displacements resulted in optimisation to a true minimum (the one shown in the table). (Using the other set of displacements resulted in another, similar structure with one imaginary frequency, not shown.) The most strongly bound structure has τ = 288° and $\Delta E_0{}^{CP}$ = − 7.6 kcal/mol, *i.e.* it is geometrically similar to the gas-phase and aqueous global minima.

Four distinct transition states in 1,4-dioxane were found. However, one of these, located by starting from τ_{twist} = 245°, was not stacked, but L-shaped (angle between planes = 98°), with a possible N-H•••Br interaction involving the N1-H1 group of uracil. "Following" the imaginary frequency of this TS yielded, for one set of displacements, the already-discovered energy minimum structure with τ_{twist} = 288°, and for the other set of displacements, a structure which was planar and appeared to contain a bifurcated hydrogen bond involving bromine.

Table 5.8 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back U/BrU minima and transition states in 1,4-dioxane PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
5	2	3	3.14	1.47	1	−2.3	−2.1	
90	74	13	3.14	0.52	0	−5.7	−5.3	−4.7
120	95	6	3.13	0.67	1	−5.5	−5.2	
165	183	6	3.08	1.58	0	−7.7	−7.3	−6.7
185	221	2	3.17	1.07	1	−6.0	−5.7	
245[a]	355	98	3.36	4.50	1	−5.7	−5.3	
205 & 290	288	9	3.04	1.03	0	−7.9	−7.6	−7.0

[a] L-shaped transition state.

Fig. 5.5 shows the gas phase interaction energy of the face-to-face U/BrU stacked dimer as a function of the τ_{twist} angle. The interaction energies are negative throughout. The BSSE is again fairly large, varying between 2.92 kcal/mol (when τ_{twist} = 45°) and 3.58 kcal/mol (when τ_{twist} = 270°). Energy minima are located at τ_{twist} = 55°, 210° and 290°. Energy maxima are located at τ_{twist} = 130°, 235° and 355°.

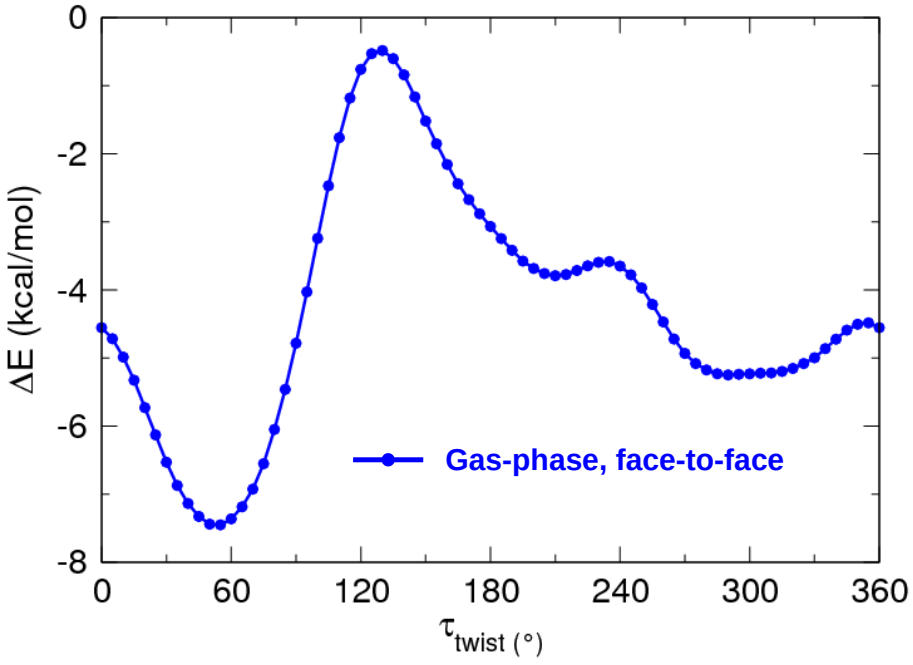


Figure 5.5: Interaction energies as a function of the twist angle for the gas-phase face-to-face U/BrU stacked dimer calculated at the M06-2X/6-31+G(d) level of theory.

Fig. 5.6 shows the interaction energies of the face-to-face U/BrU stacked dimer in PCM solvents as a function of the τ_{twist} angle. All energies are negative throughout. In water there are four minima, located at τ_{twist} = 80°, 210°, 270° and 330°, and four maxima, at 5°, 120°, 245° and 290°. In 1,4-dioxane there are also four minima, at 50°, 215°, 290° and 325°, and four maxima, at 130°, 240°, 295° and 355°. The water curve is again much flatter than the gas and 1,4-dioxane curves.

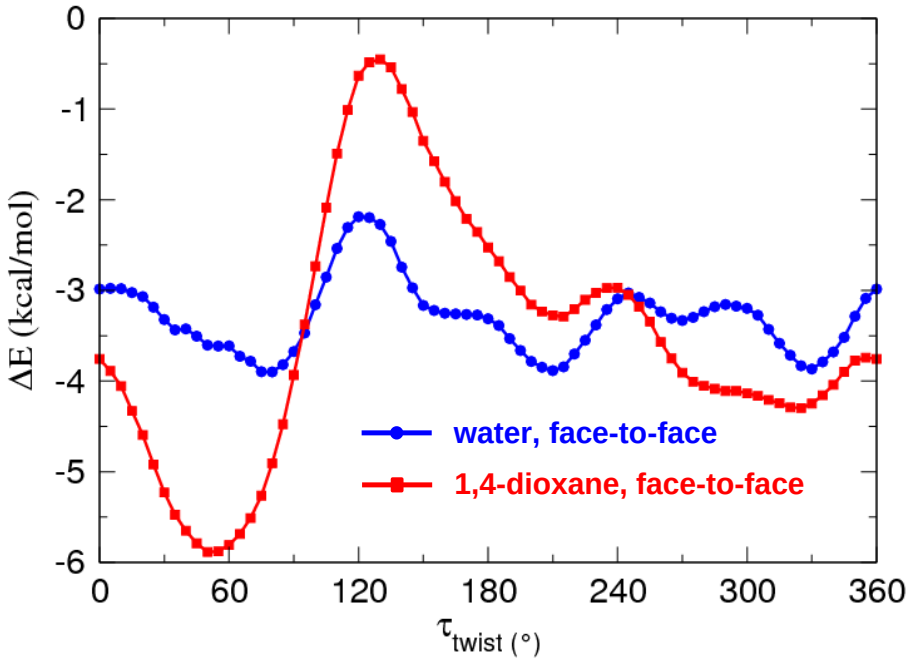


Figure 5.6: Interaction energies as a function of the twist angle for the face-to-face U/BrU stacked dimer in PCM solvent (water and 1,4-dioxane) calculated at the M06-2X/6-31+G(d) level of theory.

For the face-to-face gas-phase structures, only two distinct energy minima could be located, because the optimisations starting from τ_{twist} = 210° and τ_{twist} = 290° yielded the same structure, with a twist angle of 296° and an interaction energy of −8.5 kcal/mol. The other minimum, with an optimised twist angle of 60°, has the most negative CP-corrected, ZPE-corrected M06-2X/6-31+G(d) interaction energy of any structure we found for U/BrU (−10.7 kcal/mol). Because only two minima were located, we were only able to perform two QST3 transition state searches (starting from τ_{twist} = 130° and 355°), both of which were successful. However, a standard transition state search starting from τ_{twist} = 235° yielded an L-shaped TS with an N3H3(BrU)•••O4(U) hydrogen bond (angle between planes = 71°). "Following" the imaginary frequency yielded, for one set of atomic displacements, a planar H-bonded energy minimum structure (in non-Watson–Crick geometry), and for the other set of displacements, another H-bonded, L-shaped structure (angle between planes = 65°), this one an energy minimum.

As with the face-to-back structures, the value of the vertical separation decreased from 3.4 Å to around 3.1 Å during all the optimisations and transition state searches that yielded stacked face-to-face structures, *i.e.* the two bases moved closer together. All the energies and geometries of the gas-phase face-to-face structures are listed in Table 5.9.

Table 5.9 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face U/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
55	60	12	3.04	0.87	0	−11.5	−10.7	−5.8
130	98	14	3.14	2.36	1	−6.5	−6.1	
235[a]	n/a	71	3.15	3.73	1	−6.2	−5.7	
235[b]	n/a	65	n/a	n/a	0	−7.6	−7.0	−5.1
235[c]	n/a	n/a	n/a	n/a	0	−12.1	−11.3	−8.8
210 & 290	296	18	3.11	1.23	0	−9.1	−8.5	−3.4
355	315	10	3.10	2.23	1	−7.8	−7.3	

[a] L-shaped transition state

[b] L-shaped minimum, found by "following" the imaginary frequency of the transition state listed above it.

[c] Planar H-bonded minimum, found by "following" the imaginary frequency of the transition state listed above it.

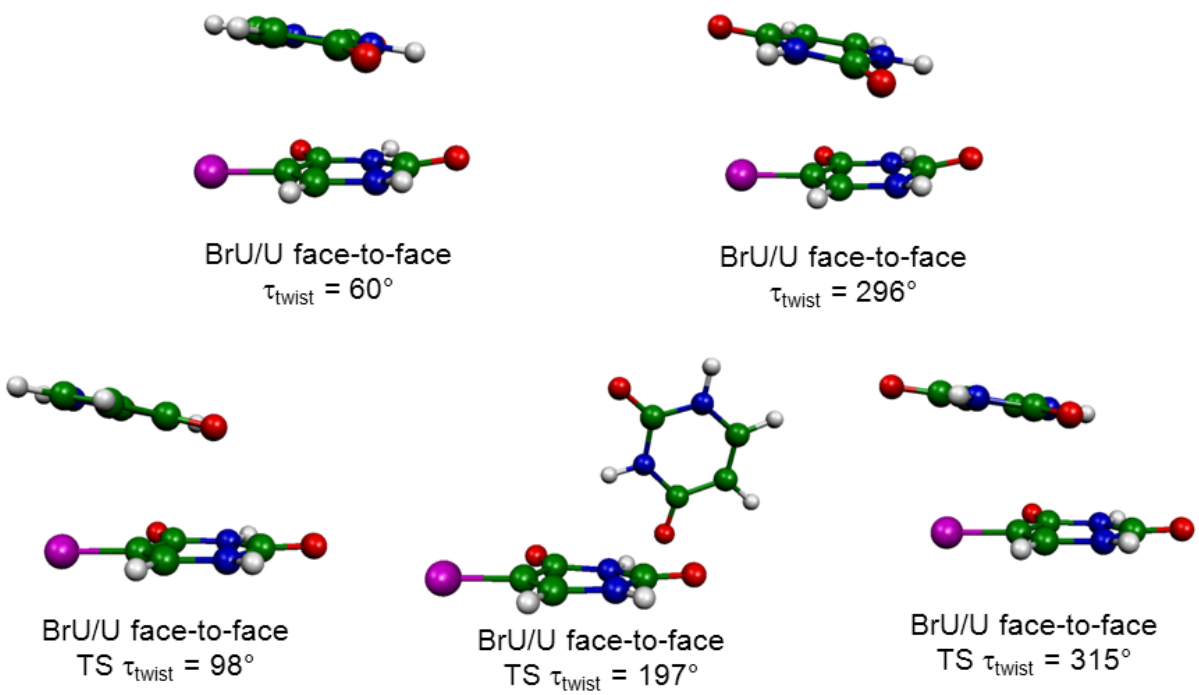


Figure 5.7: Structures of the stacked face-to-face U/BrU minima (top) and transition states (bottom) optimised in the gas phase at the M06-2X/6-31+G(d) level of theory.

For the face-to-face structures in water, four energy minima and four transition states were obtained, as listed in Table 5.10. As in the face-to-back case, the interaction energies of the optimised face-to-face structures are smaller in the two solvents than in the gas phase. The most strongly bound structure has $\tau = 71°$ and $\Delta E_0^{CP} = -5.2$ kcal/mol.

Table 5.10 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face U/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
5	43	10	3.10	1.09	1	−4.8	−4.6	
80	71	6	3.08	1.14	0	−5.6	−5.2	−4.6
120	119	1	3.16	1.31	1	−4.3	−4.0	
210	181	6	3.19	1.28	0	−4.5	−4.0	−3.4
245	250	5	3.21	1.17	1	−3.9	−3.6	
270	254	3	3.14	1.30	0	−4.8	−4.2	−3.6
295	336	3	3.17	1.13	1	−4.2	−4.0	
330	17	2	3.10	1.10	0	−5.1	−4.6	−4.0

For the face-to-face structures in 1,4-dioxane, three distinct energy minima were located from the first round of geometry optimisations, as the optimisation starting from τ_{twist} = 325° yielded the same minimum as that starting from 290°, with an optimised twist angle of 298°. Three transition states were located. However, two of these (obtained by starting from τ_{twist} = 130° and 240°) have slightly greater stabilisation energies (−4.02 kcal/mol and −4.12 kcal/mol respectively) than the minimum obtained from τ_{twist} = 215° (−3.99 kcal/mol), though they would both be expected to connect this minimum with another and therefore lie energetically above it. Furthermore, the optimised twist angle of the TS obtained from τ_{twist} = 240° is 196°, which does not lie between those of the optimised minima it would be expected to connect (209° and 298°).

For the TS with optimised τ_{twist} = 165°, "following" the imaginary frequency yielded two new stacked energy minima with optimised twist angles of 136° and 178°, the former having a large horizontal separation (2.96 Å). Both of these minima are more stable than the respective TS. However, for the TS with optimised τ_{twist} = 196°, the same procedure yielded the minimum with τ_{twist} = 178° and the already-discovered minimum with τ_{twist} = 209°. The latter structure has a very slightly greater stabilisation energy than the TS (by 0.01 kcal/mol) before the zero-point and counterpoise corrections are applied, but these corrections actually reverse the order of stability, such that the TS with τ_{twist} = 196° ($\Delta E_0{}^{CP}$ = −4.1 kcal/mol) is 0.1 kcal/mol more stable than the minimum with τ_{twist} = 209° ($\Delta E_0{}^{CP}$ = −4.0 kcal/mol), even though the attempt to optimise the TS along the appropriate reaction coordinate led to that minimum. Clearly, the use of PCM renders this area of the potential energy surface very flat, leading to complications in the identification of turning points, which is exacerbated by the necessity to apply both the energetic corrections (zero-point and counterpoise) to the optimised energies *a posteriori*. In addition, the counterpoise correction is not entirely accurate as it cannot be applied in combination with the PCM method.

The most strongly bound structure has τ = 63° and $\Delta E_0{}^{CP}$ = −8.0 kcal/mol.

Table 5.11 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face U/BrU minima and transition states in 1,4-dioxane PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
50	63	10	3.03	1.04	0	−8.6	−8.0	−7.5
130	165	14	3.28	1.72	1	−4.1	−4.0	
130[a]	178	10	3.19	1.24	0	−4.5	−4.2	−3.5
130[a]	136	21	3.04	2.96	0	−6.4	−6.0	−5.3
215	209	9	3.19	0.30	0	−4.1	−4.0	−3.2
240	196	12	3.22	0.34	1	−4.2	−4.1	
295	214	18	3.16	2.61	1	-2.5	-2.2	
290 & 325	298	11	3.07	1.48	0	−6.7	−6.3	−5.7
355	341	2	3.15	1.27	1	−5.1	−4.9	

[a] These minima were obtained by "following" the imaginary frequency of the transition state listed above them

Table 5.12 shows the RI-mPW2PLYP-D/aug-cc-pVTZ interaction energies of the gas-phase structures, using the geometries obtained from M06-2X/6-31+G(d), as well as those re-optimised with RI-mPW2PLYP-D/aug-cc-pVDZ. These were calculated to check that the M06-2X results did not significantly change at a higher level of theory. Re-optimisation of the face-to-face minimum with τ_{twist} = 296° resulted in its rearrangement into a planar H-bonded structure. The other five minima remained stacked, and only underwent slight geometrical changes. The interaction energies for those five minima always agree to within 1.0 kcal/mol with the corresponding M06-2X/6-31+G(d) energies. The ordering of the two lowest-energy face-to-back structures is changed by RI-mPW2PLYP-D: the minimum with τ_{twist} = 184° is now the most stable. However, the difference is negligible, being less than 0.2 kcal/mol even for the re-optimised structures.

Note that the RI-mPW2PLYP-D calculations included corrections for BSSE but not for deformation energies or zero-point vibrations (since the energies were not calculated at the same level of theory as the optimisations). With M06-2X, the ZPE corrections to the interaction energies had ranged from 0.6 to 0.8 kcal/mol in the face-to-back structures and from 0.7 to 0.9 kcal/mol in the face-to-face structures (these are not the ZPE corrections to the total energies but to the energy of interaction, *i.e.* $E_{ZPE}^{A/B}$ (A/B) − E_{ZPE}^{A} (A) − E_{ZPE}^{B} (B)).

Table 5.12 RI-mPW2PLYP-D/aug-cc-pVTZ interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the six gas-phase U/BrU minima identified above.

ΔE^{CP} (RI-mPW2PLYP-D)		Twist	Angle between planes	Vertical separation	Horizontal separation
M06-2X geom.[a]	re-optimised[b]				
face-to-back minima					
−9.9 (−10.1)	−10.0	184 (184)	10 (9)	3.10 (3.09)	1.51 (1.51)
−9.7 (−10.6)	−9.8	287 (287)	12 (12)	3.09 (3.06)	0.87 (0.89)
−7.9 (−7.9)	−8.1	67 (68)	22 (20)	3.24 (3.12)	0.68 (0.91)
−8.5 (−8.9)	−8.6	183 (183)	2 (2)	3.05 (3.00)	2.34 (2.37)
face-to-face minima					
−9.1 (−9.1)	−22.0[c]	n/a (296)	n/a (18)	n/a (3.11)	n/a (1.23)
−11.1 (−11.5)	−11.2	59 (60)	12 (12)	3.08 (3.04)	0.76 (0.87)

In parentheses are the original counterpoise-corrected, non-zero-point-corrected energies and optimised structural parameters from M06-2X/6-31+G(d) geometry optimisations.

[a] M06-2X/6-31+G(d)-optimised geometries.

[b] RI-mPW2PLYP-D/aug-cc-pVDZ-optimised geometries.

[c] Planar H-bonded.

The above U/BrU stacking results were published in *Theor. Chem. Acc.* (2014), **133**, p.1431.[211]

For the seven other dimers, we will not describe the specific issues encountered in the scan and optimisation procedures in full here, to save the reader time. Those details – which are generally similar to those for U/BrU – can be found in Appendix C. The potential energy graphs from the fixed scans have also been placed in Appendix C. However, we do show here the tables of geometries and interaction energies of the minima, as these are critical for the subsequent discussion.

Note that for the remaining dimers, we did not perform any calculations in 1,4-dioxane. We did calculate RI-mPW2PLYP-D/aTZ single-point interaction energies of selected gas-phase minima, and found them to be very close to the M06-2X values, as with U/BrU. Those results are not reported.

The results for the remaining dimers are currently in preparation to be submitted for publication.

5.3.2. BrU/BrU

Table 5.13 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back BrU/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	328	14	3.12	2.66	1	−5.6	−5.2	
85	73	12	3.05	0.94	0	−10.2	−9.5	−1.5
125	134	4	3.18	0.88	1	−8.6	−8.2	
145 & 215	180	0	3.15	1.90	0	−10.8	−10.2	−0.5
235	226	4	3.18	0.88	1	−8.6	−8.2	
275	287	12	3.05	0.94	0	−10.2	−9.5	−1.5

Table 5.14 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back BrU/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	25	7	3.21	1.20	1	−4.0	−4.1	
50	60	4	3.09	1.10	0	−5.6	−5.4	−4.7
65	87	4	3.15	0.80	1	−5.0	−5.1	
90	100	3	3.15	1.28	0	−5.1	−5.0	−4.2
120	87	6	3.17	1.87	1	−4.6	−4.7	
150	117	2	3.15	1.24	0	−5.4	−5.3	−4.5
185	121	7	3.17	1.70	1	−5.1	−5.1	
220	243	2	3.15	1.24	0	−5.4	−5.3	−4.5
270	260	3	3.15	1.28	0	−5.1	−5.0	−4.2
245 & 300	273	4	3.15	0.80	1	−5.0	−5.1	
315	300	4	3.09	1.10	0	−5.6	−5.4	−4.7

Table 5.15 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face BrU/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
50	49	10	3.04	0.79	0	−12.4	−11.8	−4.3
125[a]	n/a	54	n/a	n/a	1	−5.5	−5.2	
265 & 315	295	15	3.07	1.32	0	−8.8	−8.3	−0.4
355	306	8	3.10	2.11	1	−7.8	−7.5	

[a] L-shaped transition state.

Table 5.16 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face BrU/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	336	1	3.19	0.91	1	−3.7	−3.9	
45	56	4	3.10	1.08	0	−5.6	−5.6	−4.8
120 & 175	120	6	3.13	1.48	1	−5.0	−5.0	
120[a]	113	12	3.03	2.66	0	−6.0	−5.9	−5.1
150	121	5	3.13	1.38	0	−4.9	−4.8	−4.1
210	238	2	3.15	1.23	0	−5.0	−4.9	−4.1
245	220	4	3.21	1.86	1	−4.1	−4.1	
260	245	2	3.16	1.28	0	−4.9	−4.8	−4.0
290	268	4	3.17	1.12	1	−4.1	−4.1	
325	298	5	3.10	1.51	0	−4.8	−4.8	−4.0

[a] This minimum was obtained by "following" the imaginary frequency of the transition state listed above it

5.3.3. C/BrU

Table 5.17 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	1	14	3.25	1.27	1	−6.7	−5.9	
40	52	15	3.13	0.99	0	−8.5	−7.6	−3.8
120	78	6	3.03	1.80	1	−7.5	−6.8	
260	261	18	3.08	0.95	0	−15.2	−14.0	−9.1

Table 5.18 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	329	12	3.07	1.45	1	−6.0	−5.4	
0[a]	7	3	3.08	1.56	0	−5.8	−5.1	−5.0
0[a]	67	10	3.04	1.16	0	−7.1	−6.4	−5.9
40 & 85	54	9	3.16	0.67	0	−5.2	−4.7	−4.1
120	79	8	3.17	0.56	1	−5.2	−4.9	
120[b]	64	7	3.14	1.27	0	−4.8	−4.3	−3.8
120[b]	88	10	3.13	0.91	0	−5.4	−4.8	−4.2
155	119	1	3.16	1.22	0	−5.0	−4.4	−3.8
175	153	3	3.19	1.07	1	−4.5	−4.2	
210	241	5	3.16	1.16	0	−5.6	−5.0	−4.4
240	242	5	3.17	1.00	1	−5.5	−5.0	
285	293	10	3.04	1.16	0	−7.1	−6.4	−5.9

[a, b] These minima were obtained by "following" the imaginary frequencies of the transition states listed above them

Table 5.19 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
50	63	10	3.07	1.03	0	−10.4	−9.3	−4.3
105	116	2	3.13	1.12	1	−9.1	−8.3	
200	201	16	3.03	1.53	0	−14.9	−13.4	−10.1
355	63	20	3.08	2.18	1	−8.2	−7.3	

Table 5.20 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	4	3	3.17	1.28	1	−5.2	−4.6	
45	55	8	3.07	1.13	0	−6.6	−5.9	−5.5
115	127	1	3.13	1.13	1	−5.3	−4.8	
155	157	7	3.13	1.11	0	−5.4	−5.0	−4.3
185	163	8	3.12	1.17	1	−5.5	−5.2	
185[a]	187	10	3.09	1.42	0	−5.5	−4.8	−4.3
210	214	6	3.14	1.00	0	−4.9	−4.5	−3.8
240	274	8	3.19	0.49	1	−5.0	−4.5	
240[a]	16	15	3.09	1.14	0	−5.7	−5.1	−4.5
275	300	5	3.19	1.33	0	−4.8	−4.1	−3.5
310	325	5	3.24	0.97	1	−4.4	−4.0	
325	357	3	3.17	1.32	0	−5.2	−4.7	−4.1

[a] These minima were obtained by "following" the imaginary frequencies of the transition states listed above them

5.3.4. T/BrU

Table 5.21 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back T/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
5	32	16	3.13	2.60	1	−4.1	−3.7	
85	102	4	3.04	1.43	0	−8.3	−7.8	−4.0
115	130	14	3.23	1.02	1	−7.5	−7.1	
205	182	3	3.12	1.50	0	−10.1	−9.5	−3.6
240	240	4	3.19	0.96	1	−8.3	−7.8	
285	289	12	3.04	0.92	0	−11.9	−11.1	−5.5

Table 5.22 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back T/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
5	311	5	3.26	0.69	1	−4.5	−4.2	
45	58	7	3.10	0.91	0	−5.7	−5.3	−4.6
70	86	3	3.24	0.92	1	−3.8	−3.6	
90 & 150	112	3	3.16	1.21	0	−5.6	−5.2	−4.6
180	176	4	3.23	0.64	1	−5.3	−5.0	
210	223	1	3.18	0.96	0	−5.4	−5.1	−4.4
240	263	2	3.15	1.14	1	−5.7	−5.3	
285 & 315	291	6	3.07	1.10	0	−6.3	−5.9	−5.3

Table 5.23 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face T/BrU minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	28	15	3.13	2.50	1	−6.5	−6.1	
80	76	11	3.05	0.82	0	−12.0	−11.2	−6.1
130	153	9	3.12	1.96	1	−7.5	−7.0	
170	186	15	3.06	1.23	0	−10.3	−9.6	−3.7
250	211	13	3.05	1.86	1	−8.0	−7.4	
275[a]	n/a	87	n/a	n/a	0	−11.0	−10.3	−8.3

[a] L-shaped minimum.

Table 5.24 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face T/BrU minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	341	6	3.21	0.99	1	−4.7	−4.4	
45 & 80	57	6	3.08	1.15	0	−5.9	−5.7	−4.9
125	146	2	3.23	0.74	1	−3.9	−3.8	
160	187	7	3.09	1.50	0	−5.3	−4.9	−4.3
190	274	3	3.21	0.86	1	−5.2	−4.9	
190[a]	300	2	3.20	1.34	0	−5.0	−4.5	−3.9
215	225	2	3.13	1.25	0	−5.2	−4.7	−4.0
240	226	2	3.18	1.16	1	−5.1	−4.8	
275	240	3	3.15	1.24	0	−5.7	−5.2	−4.6
305	329	6	3.22	1.00	1	−4.8	−4.5	
330	344	4	3.17	1.15	0	−5.2	−4.6	−4.0

[a] This minimum was obtained by "following" the imaginary frequency of the transition state listed above it

5.3.5. C/C

Table 5.25 Interaction energy (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/C minimum in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
180	142	17	3.19	1.17	0	−12.6	−11.1	−9.6

Table 5.26 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/C minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	333	4	3.28	1.04	1	−3.3	−2.5	
80	99	17	3.15	1.08	0	−5.6	−4.4	−4.1
120	122	3	3.19	1.10	1	−4.8	−3.8	
160	141	26	3.31	1.68	0	−5.5	−4.1	−3.9
180	198	1	3.19	1.03	1	−4.8	−4.2	
200	219	26	3.31	1.68	0	−5.5	−4.1	−3.9
240	219	6	3.21	0.93	1	−4.9	−3.9	
280	261	17	3.15	1.08	0	−5.6	−4.4	−4.1

Table 5.27 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/C minima in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
210	152	20	3.23	0.71	0	−14.5	−12.6	−11.0
280[a]	n/a	n/a	n/a	n/a	0	−22.0	−20.9	−19.6

[a] Planar, hydrogen-bonded minimum.

Table 5.28 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/C minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
10	340	2	3.16	1.22	1	−4.3	−3.6	
40	20	1	3.20	1.00	0	−4.8	−4.1	−3.5
45	42	6	3.22	1.20	1	−3.9	−3.0	
80	86	7	3.23	0.45	1	−3.8	−3.1	
80	102	18	3.12	1.55	0	−5.2	−3.4	−3.6
120	158	9	3.20	1.17	1	−4.4	−3.5	
205	189	4	3.21	1.04	0	−5.2	−4.2	−3.8
245	213	5	3.23	0.69	1	−4.2	−3.3	
315	310	17	3.20	0.72	0	−5.1	−4.4	−3.8

5.3.6. C/T

Table 5.29 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/T minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
5[a]	n/a	45	n/a	n/a	1	−11.0	−10.0	
5[b]	n/a	52	n/a	n/a	0	−13.9	−12.5	−10.8
5[c]	n/a	n/a	n/a	n/a	0	−20.1	−19.5	−18.2
85	84	15	2.90	2.52	0	−8.1	−7.1	−5.3
115	149	6	3.06	1.55	1	−8.8	−7.9	
115[d]	132	4	3.00	1.69	0	−9.0	−8.0	−6.1
210	222	10	3.13	0.76	0	−12.1	−10.9	−9.2

[a] L-shaped transition state.
[b] L-shaped minimum, found by "following" the imaginary frequency of the transition state listed above it.
[c] Planar H-bonded minimum, found by "following" the imaginary frequency of the transition state listed above it.
[d] Stacked minimum, found by "following" the imaginary frequency of the transition state listed above it.

Table 5.30 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/T minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	48	6	3.20	1.48	1	−4.5	−3.7	
80	61	6	3.12	1.11	0	−5.1	−4.4	−3.8
120	123	2	3.18	1.05	0	−4.9	−4.2	−3.6
160	150	3	3.21	0.88	0	−5.2	−4.5	−3.9
170	167	1	3.25	0.05	1	−4.5	−3.9	
205	210	6	3.19	0.84	0	−5.6	−4.8	−4.2
240	230	7	3.19	1.03	1	−5.6	−4.8	
240[a]	224	8	3.19	1.03	0	−5.6	−4.6	−4.1
240[a]	250	6	3.18	1.22	0	−5.8	−4.8	−4.4
285 & 315	296	12	3.07	1.11	0	−6.7	−5.8	−5.3

[a] These minima were obtained by "following" the imaginary frequency of the transition state listed above them

Table 5.31 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/T minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	38	16	3.11	1.14	1	−6.9	−5.8	
50	72	13	3.10	1.19	1	−8.0	−7.0	
50[a]	79	14	3.08	1.52	0	−8.1	−7.0	−5.1
105	158	14	3.06	2.42	1	−9.4	−8.6	
105[b]	188	21	3.13	1.00	0	−12.6	−11.6	−9.7
105[b, c]	n/a	80	n/a	n/a	0	−11.0	−10.1	−8.7
210	209	16	2.92	2.22	0	−13.4	−12.1	−10.3

[a, b] These minima were obtained by "following" the imaginary frequency of the transition state listed above them

[c] L-shaped minimum

Table 5.32 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/T minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	357	2	3.16	1.31	1	−5.6	−4.7	
0[a]	10	12	3.06	1.51	0	−6.0	−5.1	−4.7
0[a]	354	3	3.18	1.26	0	−5.7	−4.7	−4.4
40	61	10	3.10	1.06	0	−6.2	−5.2	−4.8
75	88	9	3.12	1.05	1	−5.9	−5.3	
85	94	7	3.12	0.99	0	−5.9	−5.1	−4.6
115	139	6	3.18	1.12	1	−4.9	−4.4	
155	178	11	3.12	1.22	0	−5.7	−4.7	−4.3
195	197	11	3.11	1.56	1	−4.8	−4.2	
210	182	2	3.13	1.21	0	−5.2	−4.4	−3.9
240	270	12	3.18	0.72	1	−5.1	−4.3	
275	288	4	3.22	0.88	0	−5.7	−4.7	−4.3

[a] These minima were obtained by "following" the imaginary frequency of the transition state listed above them

5.3.7. C/U

Table 5.33 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/U minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	356	21	3.33	1.31	1	−5.3	−4.5	
55	52	19	3.13	1.22	0	−7.0	−6.0	−4.2
115	106	7	3.14	1.44	1	−6.4	−5.5	
210	220	10	3.15	0.76	0	−11.3	−10.2	−8.5

Table 5.34 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back C/U minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
0	7	7	3.19	1.34	1	−3.5	−2.8	
55	56	14	3.22	0.40	1	−4.6	−4.0	
55[a]	58	12	3.22	0.38	1	−4.6	−4.1	
55[a]	60	6	3.13	1.21	0	−4.8	−4.1	−3.5
120	149	3	3.23	0.84	1	−4.2	−3.6	
195	216	7	3.21	0.92	0	−4.9	−4.2	−3.6
240	233	1	3.19	1.02	1	−4.7	−4.0	
295	295	13	3.08	1.16	0	−6.0	−5.0	−4.5

Table 5.35 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/U minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}
50	80	15	3.06	1.66	0	−8.1	−7.0	−5.2
105	160	4	3.02	2.11	1	−9.5	−8.6	
105[a]	143	9	3.01	2.10	0	−9.5	−8.5	−6.8
105[a,b]	n/a	n/a	n/a	n/a	0	−20.4	−19.4	−18.0
200	209	17	2.91	2.26	0	−13.9	−12.6	−10.8
355	350	21	3.26	1.81	1	−6.0	−5.0	

[a] These minima were obtained by "following" the imaginary frequency of the transition state listed above them.
[b] Planar hydrogen-bonded minimum.

Table 5.36 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face C/U minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	356	3	3.16	1.40	1	−4.5	−3.9	
0[a]	15	19	3.13	1.48	0	−5.4	−4.5	−4.0
0[a]	353	3	3.18	1.32	0	−4.8	−3.9	−3.5
40	77	18	3.10	1.43	0	−5.5	−4.5	−4.1
115	146	4	3.18	1.24	1	−4.2	−3.6	
165	172	16	3.22	0.58	0	−4.8	−4.2	−3.6
240	274	8	3.21	0.60	1	−4.4	−3.7	
285	292	5	3.22	0.86	0	−4.8	−3.8	−3.4

[a] These minima were obtained by "following" the imaginary frequency of the transition state listed above them.

5.3.8. T/U

Table 5.37 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back T/U minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	350	4	3.17	1.55	1	−3.2	−2.8	
75	71	16	3.09	0.87	0	−8.9	−8.1	−6.1
120	134	24	3.33	1.25	1	−7.3	−6.8	
185	182	6	3.09	1.23	0	−9.9	−9.1	−7.4
245	209	10	3.16	0.92	1	−7.4	−6.8	
290	289	17	3.09	0.91	0	−9.6	−8.8	−6.8

Table 5.38 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back T/U minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	332	7	3.24	1.00	1	−4.2	−3.6	
60	66	11	3.10	0.95	0	−5.6	−4.9	−4.3
120	113	4	3.18	1.00	1	−4.7	−4.1	
160	175	4	3.17	1.20	0	−5.2	−4.5	−4.0
190	198	2	3.24	0.12	1	−4.3	−3.9	
205	215	1	3.20	0.88	0	−5.1	−4.5	−3.9
240	231	2	3.19	1.05	1	−4.9	−4.3	
285	292	11	3.10	0.96	0	−5.3	−4.6	−4.0

Table 5.39 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face T/U minima and transition states in the gas phase calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	20	14	3.13	2.47	1	−5.1	−4.6	
75	75	15	3.07	0.88	0	−10.2	−9.3	−7.4
120	161	11	3.11	2.18	1	−7.4	−6.9	
185	184	19	3.16	0.90	0	−9.8	−8.8	−7.1
250	288	15	3.15	1.18	1	−6.7	−6.1	
290	298	19	3.27	0.61	0	−6.9	−6.4	−4.4

Table 5.40 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-face T/U minima and transition states in water PCM solvent calculated at the M06-2X/6-31+G(d) level of theory

τ_{twist} (initial)	τ_{twist} (final)	Angle between planes	Vertical separation	Horizontal separation	Im. Freq.	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}
0	22	6	3.11	1.41	1	−4.8	−4.1	
40	62	11	3.09	1.04	0	−5.5	−4.8	−4.3
80	81	12	3.12	0.91	1	−5.2	−4.5	
90	95	7	3.11	0.87	0	−5.4	−4.6	−4.1
115	141	3	3.22	1.01	1	−4.5	−4.0	
160	175	9	3.14	1.26	0	−5.0	−4.3	−3.8
240	264	7	3.23	0.93	1	−4.8	−4.1	
280	292	9	3.20	0.25	0	−4.9	−4.3	−3.7

5.4. Discussion

In terms of potential energy, the most strongly bound stacked Py/Py structure of any found in this study is the gas-phase, face-to-back C/BrU dimer with $\tau = 261°$, for which $\Delta E_0^{CP} = -14.0$ kcal/mol at the M06-2X/6-31+G(d) level of theory (Table 5.17). If we compare this to the interaction potential energy of the most strongly bound face-to-back C/T stacked dimer, for which $\tau = 222°$ and $\Delta E_0^{CP} = -10.9$ kcal/mol (Table 5.29), the difference is 3.1 kcal/mol in favour of BrU stacking: not a negligible amount, following the arguments of Loeb/Kunkel[19] and Lasken/Goodman[154] outlined in Chapter 3, Section 3.5. Likewise, the most strongly stacked face-to-back C/U dimer, with $\tau = 220°$, has $\Delta E_0^{CP} = -10.2$ kcal/mol (Table 5.33), i.e. the interaction is almost 4.0 kcal/mol weaker than in C/BrU.

We can make similar observations for the cases where BrU, T and U stack with either T or U. The most strongly bound face-to-back T/BrU dimer has $\tau = 289°$ and $\Delta E_0^{CP} = -11.1$ kcal/mol, while the equivalent values for T/T (from Ref. [206]) are 70°/290° and −9.4 kcal/mol, and those for T/U are 182° and −9.1 kcal/mol. The most strongly bound face-to-back U/BrU dimer has $\tau = 287°$ and $\Delta E_0^{CP} = -9.8$ kcal/mol, while the equivalent values for T/U are 182° and −9.1 kcal/mol, and those for U/U

(from Ref. [206]) are 180° and −8.5 kcal/mol. Table 5.41 lists the geometric parameters and stacking potential energies of the most strongly bound gas-phase FTB dimers for each system.

Table 5.41 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the most strongly bound face-to-back minima in the gas phase for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	ΔE_0^{CP}
C/BrU	261	18	3.08	0.95	−15.2	−14.0
C/T	222	10	3.13	0.76	−12.1	−10.9
C/U	220	10	3.15	0.76	−11.3	−10.2
T/BrU	289	12	3.04	0.92	−11.9	−11.1
T/T[a]	70 / 290	14	3.05	0.90	−10.3	−9.4
T/U	182	6	3.09	1.23	−9.9	−9.3
U/BrU	287	12	3.06	0.89	−10.6	−9.8
U/U[a]	180	0	3.08	1.23	−9.3	−8.5
BrU/BrU	180	0	3.15	1.90	−10.8	−10.2
C/C	142	17	3.19	1.17	−12.6	−11.1

[a] From Ref.[206]

At first glance, therefore, the results seem to support the hypothesis that BrU undergoes stronger stacking interactions than T and U – at least when the partner base is C, T or U. Setting the ZPE-corrected stacking "strength" of BrU as 1.00, the relative stacking strengths of T and U (i.e. the ratios of ΔE_0^{CP} to that of BrU) are as follows: for stacking with C, strength of T = 0.78 and U = 0.73; for stacking with T, strength of T = 0.85 and U = 0.84; for stacking with U, strength of T = 0.95 and U = 0.87. These follow the same trend as the gas-phase polarisabilities (Table 5.5), the ratios of which are BrU = 1.00, T = 0.92 and U = 0.77.

However, a more detailed consideration of the results challenges this conclusion.

Firstly, none of the above dimers has a value of the "twist" angle, τ, that is anywhere near the canonical B-DNA value of the Helical Twist angle, 36°. In fact, among all our gas-phase, face-to-back stacked energy minima, the optimised value of τ that comes closest to 36° is 52° (for C/BrU and C/U), while for C/T (for example) it is even more divergent, at 84°. Clearly, in the absence of either a sugar-phosphate backbone or a third stacker, stacked dimers prefer geometries that depart considerably from the canonical stacking arrangements encountered in the cell. The familiar *in vivo* stacking motifs are not true energy minima with respect to the interaction of any two neighbouring bases, but are compromises, enforced by the global requirement to maintain DNA helicity, and the local requirement for each base to stack favourably with its neighbours on both sides, while also base-pairing with the opposite strand. The structures of real intra-helical dinucleotides thus represent a balance between several competing demands, while the situation for covalently separated, non-base-paired, stacked dimers is very different.

Therefore, to assess whether BrU stacks more strongly than T and U *in vivo*, it is more instructive to restrict the comparison to those energy minima for each dimer that have τ closest to ±36°. When this is done, for stacking with C, the earlier qualitative conclusion is not changed: C/BrU (τ = 52°) has a stacking energy of −7.6 kcal/mol, which is slightly greater than C/T (τ = 84°, ΔE_0^{CP} = −7.1 kcal/mol)

and C/U (τ = 52°, ΔE_0^{CP} = −6.0 kcal/mol). The same also holds for stacking with T: T/BrU (τ = 289° ≡ −71°) has a stacking energy of −11.1 kcal/mol, which is greater than T/T (τ = 70°/290°, ΔE_0^{CP} = −9.4 kcal/mol) and T/U (τ = 289° ≡ −71°, ΔE_0^{CP} = −8.8 kcal/mol). However, for stacking with U, BrU is actually the weakest: U/BrU (τ = 68°) has a stacking energy of −7.2 kcal/mol, which is weaker than T/U (τ = 289° ≡ −71°, ΔE_0^{CP} = −8.8 kcal/mol) and even slightly weaker than U/U (τ = 72°/288°, ΔE_0^{CP} = −7.5 kcal/mol). This comparison is summarised in Table 5.42.

Table 5.42 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back minima with τ closest to ±36° in the gas phase, for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	ΔE_0^{CP}
C/BrU	52	15	3.13	0.99	−8.5	−7.6
C/T	84	15	2.90	2.52	−8.1	−7.1
C/U	52	19	3.13	1.22	−7.0	−6.0
T/BrU	289	12	3.04	0.92	−11.9	−11.1
T/T[a]	70 / 290	14	3.05	0.90	−10.3	−9.4
T/U	289	17	3.09	0.91	−9.6	−8.8
U/BrU	68	20	3.12	0.91	−7.9	−7.2
U/U[a]	72 / 288	20	3.14	0.83	−8.4	−7.5
BrU/BrU	73 / 287	12	3.05	0.94	−10.2	−9.5
C/C	142	17	3.19	1.17	−12.6	−11.1

[a] From Ref.[206]

Secondly, in water, the absolute stacking energies are smaller than in the gas phase, hence the differences between dimers also tend to be smaller. Repeating, firstly, the comparison of *all* energy minima (i.e. allowing all values of τ) yields the following results in water: for stacking with C, C/BrU has τ = 67/293° and ΔE_0^{CP} = −6.4 kcal/mol, C/T has τ = 296° and ΔE_0^{CP} = −5.8 kcal/mol, and C/U has τ = 295° and ΔE_0^{CP} = −5.0 kcal/mol; for stacking with T, T/BrU has τ = 291° and ΔE_0^{CP} = −5.9 kcal/mol, T/T has τ = 61°/299° and ΔE_0^{CP} = −5.5 kcal/mol, and T/U has τ = 66° and ΔE_0^{CP} = −4.9 kcal/mol; and for stacking with U, U/BrU has τ = 293° and ΔE_0^{CP} = −5.2 kcal/mol, T/U has τ = 66° and ΔE_0^{CP} = −4.9 kcal/mol, and U/U has τ = 66°/294° and ΔE_0^{CP} = −4.1 kcal/mol. Therefore, when stacking with C, T or U, the order of stacking strength is always BrU > T > U, as in the gas phase; however, the difference between BrU and T is always less than 1 kcal/mol. Such small differences are probably less than the error margin of this methodology. This comparison is summarised in Table 5.43.

Table 5.43 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the most strongly bound face-to-back minima in water PCM solvent for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	ΔE_0^{CP}
C/BrU	67 / 293	10	3.04	1.16	−7.1	−6.4
C/T	296	12	3.07	1.11	−6.7	−5.8
C/U	295	13	3.08	1.16	−6.0	−5.0
T/BrU	291	6	3.07	1.10	−6.3	−5.9
T/T[a]	61 / 299	10	3.09	0.91	−6.4	−5.5
T/U	66	11	3.10	0.95	−5.6	−4.9
U/BrU	293	6	3.09	1.12	−5.5	−5.2
U/U[a]	66 / 294	12	3.12	0.94	−4.7	−4.1
BrU/BrU	60 / 300	4	3.09	1.10	−5.6	−5.4
C/C	99 / 261	17	3.15	1.08	−5.6	−4.4

[a] From Ref.[206]

Then, repeating the comparison of the minima with τ closest to ±36° yields the following results: for stacking with C, C/BrU has τ = 54° and ΔE_0^{CP} = −4.7 kcal/mol, C/T has τ = 61° and ΔE_0^{CP} = −4.4 kcal/mol, and C/U has τ = 295° and ΔE_0^{CP} = −5.0 kcal/mol; for stacking with T, T/BrU has τ = 58° and ΔE_0^{CP} = −5.3 kcal/mol, T/T has τ = 61°/299° and ΔE_0^{CP} = −5.5 kcal/mol, and T/U has τ = 66° and ΔE_0^{CP} = −4.9 kcal/mol; and for stacking with U, U/BrU has τ = 293° and ΔE_0^{CP} = −5.2 kcal/mol, T/U has τ = 66° and ΔE_0^{CP} = −4.9 kcal/mol, and U/U has τ = 66°/294° and ΔE_0^{CP} = −4.1 kcal/mol. This comparison is summarised in Table 5.44. (The minimum with τ closest to ±36° is also the deepest minimum for all aqueous FTB dimers except C/BrU, C/T and T/BrU.)

No clear pattern emerges from these results, and the energy differences are in any case very small, typically less than 1 kcal/mol. Hence it cannot be argued that BrU, under aqueous conditions and with DNA-like geometries, stacks more strongly than T or U when the stacking partner is C, T or U.

Table 5.44 Interaction energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back minima with τ closest to ±36° in water PCM solvent, for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	ΔE_0^{CP}
C/BrU	54	9	3.16	0.67	−5.2	−4.7
C/T	61	6	3.12	1.11	−5.1	−4.4
C/U	295	13	3.08	1.16	−6.0	−5.0
T/BrU	58	7	3.10	0.91	−5.7	−5.3
T/T[a]	61 / 299	10	3.09	0.91	−6.4	−5.5
T/U	66	11	3.10	0.95	−5.6	−4.9
U/BrU	293	6	3.09	1.12	−5.5	−5.2
U/U[a]	66 / 294	12	3.12	0.94	−4.7	−4.1
BrU/BrU	60 / 300	4	3.09	1.10	−5.6	−5.4
C/C	99 / 261	17	3.15	1.08	−5.6	−4.4

[a] From Ref. [206]

Note that face-to-face stacking does not occur within a single DNA strand, so the FTF results are not directly applicable to the model of BrU-mutagenicity driven by enhanced stacking. However, it is interesting to note that the stacking strengths of the most strongly bound face-to-face C/X, T/X and U/X dimers (X = BrU, T or U) are always in the order BrU > T > U, in both the gas phase and aqueous solution. The full comparison will not be given here, but the differences between the optimum stacking energies for X = BrU and X = T (denoted $\Delta\Delta E_0^{CP}$) are as follows: for gas-phase C/X, $\Delta\Delta E_0^{CP}$ = 1.7 kcal/mol; for aqueous C/X, $\Delta\Delta E_0^{CP}$ = 0.7 kcal/mol; for gas-phase T/X, $\Delta\Delta E_0^{CP}$ = 1.2 kcal/mol; for aqueous T/X, $\Delta\Delta E_0^{CP}$ = 0.1 kcal/mol; for gas-phase U/X, $\Delta\Delta E_0^{CP}$ = 1.8 kcal/mol; and for aqueous U/X, $\Delta\Delta E_0^{CP}$ = 1.3 kcal/mol (always in favour of X = BrU).

Thirdly, stacking in face-to-back BrU/BrU is weaker than in T/BrU, even in the gas phase: when all values of τ are considered, the BrU/BrU global minimum has τ = 180° and ΔE_0^{CP} = −10.2 kcal/mol, while T/BrU has τ = 289° and ΔE_0^{CP} = −11.1 kcal/mol; and when only the τ values closest to ±36° are included, BrU/BrU has τ = 73°/287° and ΔE_0^{CP} = −9.5 kcal/mol, while T/BrU has τ = 289° and ΔE_0^{CP} = −11.1 kcal/mol. (In aqueous solution, the corresponding differences are less than 0.5 kcal/mol, and can be considered negligible.) Although BrU is not a natural component of DNA, it can be artificially induced to replace almost 100% of T residues in DNA,[59] so BrU/BrU stacked dimers can occur *in vitro*.

Fourthly, thermal effects reduce the stacking strength of BrU (relative to T and U) even further, and actually nullify the preference for BrU. When the enthalpic corrections at 298.15 K are added, the strengths of all the stacking interactions decrease compared to the ZPE-corrected potential energies, but this effect is greatest for dimers containing BrU, especially in the gas phase. Let us focus on face-to-back, gas-phase stacking. For U/BrU, the interaction of the most strongly bound dimer is reduced from −9.8 to −4.7 kcal/mol. For BrU/BrU, the decrease is even sharper, from −10.2 to −0.5 kcal/mol. For C/BrU it is from −14.0 to −9.1 kcal/mol, and for T/BrU it is from −11.1 to −5.5 kcal/mol. In comparison, the corresponding decrease for C/T is from −10.9 to −9.2 kcal/mol, for C/U is from −10.2 to −8.5 kcal/mol, and for T/U is from −9.1 to −7.4 kcal/mol. Using unpublished data[212] from the study by Hunter and van Mourik (Ref.[206]), we have calculated that the corresponding decrease for T/T is from −9.4 to −7.5 kcal/mol, and for U/U is from −8.5 to −6.9 kcal/mol. Therefore, when all values of τ are allowed, the three bases under comparison stack in the order of strength T > U > BrU when the stacking partner is T, U or BrU, and in the order T > BrU > U when the stacking partner is C. This comparison is summarised in Table 5.45.

Note that the most strongly bound BrU/BrU dimer in terms of potential energy (with τ = 180°; see Table 5.41) is not the most strongly bound in terms of enthalpy (with τ = 73° or 287°).

Table 5.45 Interaction enthalpies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the most strongly bound face-to-back minima in the gas phase for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔH^{CP}
C/BrU	261	18	3.08	0.95	−9.1
C/T	222	10	3.13	0.76	−9.2
C/U	220	10	3.15	0.76	−8.5
T/BrU	289	12	3.04	0.92	−5.5
T/T[a]	180	0	3.12	1.22	−7.6
T/U	182	6	3.09	1.23	−7.4
U/BrU	287	12	3.06	0.89	−4.7
U/U[a]	180	0	3.08	1.23	−6.9
BrU/BrU	73 / 287	12	3.05	0.94	−1.5
C/C	142	17	3.19	1.17	−9.6

[a] Unpublished data[212] from the authors of Ref.[206]

When only the values of τ closest to ±36° are chosen, the order is T > U > BrU for all four stacking partners. In other words, in enthalpic terms, BrU takes part in *weaker* stacking interactions with pyrimidines than either T or U does. This comparison is summarised in Table 5.46.

Table 5.46 Interaction enthalpies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back minima with τ closest to ±36° in the gas phase, for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔH^{CP}
C/BrU	52	15	3.13	0.99	−3.8
C/T	84	15	2.90	2.52	−5.3
C/U	52	19	3.13	1.22	−4.2
T/BrU	289	12	3.04	0.92	−5.5
T/T[a]	70 / 290	14	3.05	0.90	−7.5
T/U	289	17	3.09	0.91	−6.8
U/BrU	68	20	3.12	0.91	−3.4
U/U[a]	72 / 288	20	3.14	0.83	−5.6
BrU/BrU	73 / 287	12	3.05	0.94	−1.5
C/C	142	17	3.19	1.17	−9.6

[a] Unpublished data[212] from the authors of Ref.[206]

The equivalent comparison in water is shown in Table 5.47. Like the aqueous potential energies, the aqueous enthalpies show very little variation among dimers – BrU stacks more strongly than T when the partner is C, U or BrU, while the reverse is true when the partner is T, but in each case the difference between T and BrU is less than 1 kcal/mol. Therefore, in water, T and BrU appear to undergo essentially the same enthalpy of stacking with pyrimidines.

Table 5.47 Interaction enthalpies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of the face-to-back minima with τ closest to ±36° in water PCM solvent, for each system, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	Angle between planes	Vertical separation	Horizontal separation	ΔH^{CP}
C/BrU	54	9	3.16	0.67	−4.1
C/T	61	6	3.12	1.11	−3.8
C/U	60	6	3.13	1.21	−3.5
T/BrU	58	7	3.10	0.91	−4.6
T/T[a]	61 / 299	10	3.09	0.91	−5.1
T/U	66	11	3.10	0.95	−4.3
U/BrU	293	6	3.09	1.12	−4.5
U/U[a]	66 / 294	12	3.12	0.94	−3.5
BrU/BrU	60 / 300	4	3.09	1.10	−4.7
C/C	99 / 261	17	3.15	1.08	−4.1

[a] Unpublished data[212] from the authors of Ref.[206]

When the entropic corrections are included, the stacking of all the base dimers is actually unfavourable in terms of free energy (unlike the formation of hydrogen-bonded base pairs, which is usually favourable). Table 5.48 shows the counterpoise-corrected Gibbs free energies of interaction, ΔG^{CP}, of a representative sample of the optimised gas-phase dimers described above. All the stacked dimers have positive (repulsive) interaction free energies, not only in the gas phase, but also in water. However, the planar, hydrogen-bonded C/T gas-phase structure, found by "following" the frequency of the L-shaped transition state (third row in Table 5.29), has a negative value of ΔG^{CP}, showing that the free energy correction does not disfavour base *pairing* nearly as strongly as it disfavours stacking.

Table 5.48 Gibbs free energies of interaction (in kcal/mol) of selected face-to-back minima in the gas phase or water PCM solvent, calculated at the M06-2X/6-31+G(d) level of theory

System	τ_{twist}	ΔG^{CP}
C/BrU, gas	52	+7.4
C/T, gas	84	+6.0
C/T, gas[a]	n/a	−6.9
C/U, gas	52	+6.7
T/BrU, gas	289	+6.6
T/U, water	292	+6.6
U/BrU, water	67	+7.5
BrU/BrU, water	60 / 300	+7.0
C/C, water	141	+7.3

[a] Planar hydrogen-bonded minimum

In summary, although the gas-phase potential energies show enhanced stacking of BrU, this finding does not withstand the consideration of three additional factors: similarity to DNA-like geometry, influence of solvent, and thermal effects. When the comparison of gaseous *enthalpies* is restricted to dimers with DNA-like conformations, BrU is actually a weaker stacker than T or U. In water, meanwhile, the differences between the stacking of BrU, T and U are negligible in both enthalpy and potential energy. On the basis of these results, it seems unlikely that the enhanced mutagenicity of BrU can be attributed to better stacking with pyrimidines.

How important is the influence of solvent on stacking? Kabeláč and Hobza found, using classical molecular dynamics, that for the stacking of a base dimer to be thermodynamically competitive with base pairing at non-zero temperatures, between two and eight explicit water molecules were required (depending on the identity of the bases).[14] Experimentally, Nakata *et al.* observed stacking-like behaviour (the formation of columnar liquid crystal phases in solution) in DNA duplexes as short as six base pairs in length.[213] However, our calculations did not include explicit water molecules.

Why was stacking found to be entropically unfavourable, in both the gas phase and water? Stacking is known to occur in water, so if it were driven by free energy, the repulsive stacking energies in our aqueous results would suggest that the continuum model neglects the hydrophobic interactions that are responsible for aqueous stacking. This in turn would imply that the difference between gas-phase and aqueous stacking was not simply the presence of the solvent's electric field. That would be consistent with the findings of Danilov *et al*., who showed, using Monte Carlo simulations, that base stacking in water was primarily driven by favourable base–water interactions, which can only be revealed when the solvent molecules are explicitly modelled and are allowed to move.[17] The authors suggested that this favourable interaction was due to the smaller hydrophobic surface area of stacked dimers compared to hydrogen-bonded base pairs.

However, Danilov *et al.* also cited previous studies showing that base stacking in solution was primarily driven by *enthalpy*, rather than free energy.[17] Furthermore, we found stacking to be entropically disfavoured in the gas phase as well as water. Hence there are two possible explanations for our positive ΔG values: they may be due to inaccuracies in the calculation of S_{vib}, affecting both phases, perhaps resulting from the harmonic approximation; or it may be that the stacking of individual base dimers really is entropically disfavoured at room temperature.

5.5. Addendum: Py/Pu geometry optimisations

We have not attempted a complete search of the stacking potential energy surfaces of the pyrimidines, T, U and BrU, with the purines, A and G, due to the greater complexity introduced by the larger surface areas of purines. However, we did wish to test whether the general conclusions for Py/Py stacking in the context of BrU-mutagenicity – that in DNA-like conformations, BrU stacks more strongly than T in gaseous potential energy terms, but not in aqueous enthalpic terms – held for Py/Pu stacking. We therefore optimised a selection of the experimental Py/Pu dimers previously studied in Section 5.2, in both the gas phase and water PCM.

When Py = BrU or T, and Pu = A or G, there are eight possible Py/Pu combinations in DNA, taking into account strand directionality: A/BrU, BrU/A, A/T, T/A, G/BrU, BrU/G, G/T and T/G. We have chosen two examples of each, and optimised them using the same model chemistries as for the Py/Py dimers in Section 5.3. The resulting geometries and interaction energies are shown for the gas phase

in Table 5.49, and for water in Table 5.50. As can be seen, in some cases, two starting structures from different PDB entries converged to the same optimised geometry.

The gas-phase T/A dimer from 2L2V optimised to an L-shaped rather than stacked structure. The gas-phase G/T dimer from 3G6X initially failed to reach a minimum, but two different gas-phase minima for this structure were located as follows: one by re-optimising the corresponding PCM minimum in the gas phase, the other by first optimising the starting structure at the ωB97x-D/6-31+G(d) level of theory,[214] then re-optimising this minimum with M06-2X. The same problem was encountered with the BrU/G dimer from 1DA1; in this case, re-optimising the PCM minimum yielded a stacked minimum, while re-optimisation with ωB97x-D yielded a planar, H-bonded minimum (not shown in the table).

In water, the A/BrU dimers from 1A35 and 1P1Y both optimised to structures with one imaginary frequency, even when the optimisations were re-started from the corresponding PCM or ωB97x-D minima. The same occurred when we tried dimers from 1DCR and 1GQU. "Following" the imaginary frequencies also failed to locate any true minima. The aqueous A/BrU dimers are therefore absent from Table 5.50.

Table 5.49 Interaction potential energies, enthalpies and free energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of face-to-back Py/Pu minima, optimised from PDB starting structures, in the gas phase at the M06-2X/6-31+G(d) level of theory

Dimer	PDB code	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	ΔE_0^{CP}	ΔH^{CP}	ΔG^{CP}
A/BrU	1A35, 1P1Y	7	3.18	0.50	−8.5	−7.4	−2.1	+9.0
A/T	1BNA, 1A35	7	3.19	0.44	−8.5	−7.2	−5.0	+6.3
BrU/A	366D	32	2.78	3.71	−8.1	−7.0	−4.0	+5.1
BrU/A	1FN1	2	3.13	0.83	−8.9	−7.8	−3.0	+8.5
T/A	2ORF	6	3.15	1.97	−6.9	−6.0	−3.6	+6.5
T/A	2L2V[a]	51	2.99	2.89	−8.2	−6.9	−5.0	+4.9
G/BrU	1FN1	9	3.19	0.71	−10.0	−9.4	−4.4	+6.7
G/BrU	344D	10	3.00	3.68	−7.9	−7.4	−2.6	+7.8
G/T	3G6X[b]	3	3.02	1.66	−16.2	−14.9	−12.7	−0.2
G/T	3G6X[c], 2L5K	8	3.03	2.30	−12.2	−11.1	−9.0	+2.6
BrU/G	1D9R	4	3.02	1.58	−17.0	−15.9	−11.2	+1.3
BrU/G	1DA1[b]	1	3.07	1.02	−16.4	−15.5	−9.2	+2.9
T/G	3OH9, 1CS7	3	3.02	1.66	−16.2	−14.9	−12.7	−0.2

[a] L-shaped

[b] Optimised from the corresponding water PCM minimum

[c] Optimised from the corresponding ωB97x-D/6-31+G(d) minimum

Table 5.50 Interaction potential energies, enthalpies and free energies (in kcal/mol) and structural parameters (distances in Å, angles in degrees) of face-to-back Py/Pu minima, optimised from PDB starting structures, in water PCM at the M06-2X/6-31+G(d) level of theory

Dimer	PDB code	Angle between planes	Vertical separation	Horizontal separation	ΔE^{CP}	$\Delta E_0{}^{CP}$	ΔH^{CP}	ΔG^{CP}
A/BrU	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
A/T	1BNA, 1A35	3	3.16	0.55	−6.8	−5.7	−5.4	+6.7
BrU/A	366D	1	3.19	0.53	−6.2	−5.6	−5.1	+6.7
BrU/A	1FN1	19	3.11	1.72	−5.6	−5.0	−4.4	+6.3
T/A	2ORF	3	3.14	0.74	−6.6	−5.7	−5.3	+6.5
T/A	2L2V	23	3.08	2.17	−5.6	−4.7	−4.3	+6.7
G/BrU	1FN1	8	3.14	0.83	−6.5	−5.9	−5.3	+6.3
G/BrU	344D	6	3.19	2.54	−5.7	−5.4	−4.6	+6.5
G/T	3G6X	8	3.15	0.82	−7.2	−6.1	−5.7	+6.2
G/T	2L5K	8	3.16	0.84	−7.2	−6.2	−5.8	+6.1
BrU/G	1D9R	6	3.10	1.94	−6.7	−6.0	−5.4	+6.2
BrU/G	1DA1	6	3.16	0.98	−6.1	−5.7	−5.0	+5.7
T/G	3OH9, 1CS7	6	3.13	0.70	−7.1	−6.2	−5.7	+6.0

When the partner is adenine, in the gas phase, BrU and T stack about equally strongly in terms of potential energy, but in enthalpic terms, the advantage is slightly with T. In water, the differences are negligible for both energies.

When the partner is guanine, in the gas phase, T has a clear advantage in potential energy terms, but only when it is on the 3' side. In gas-phase enthalpic terms, T stacks more strongly than BrU, whichever side it is on. In water, the differences are again negligible for both energies.

The results therefore reinforce the conclusions from Sections 5.2 and 5.3: no evidence was found that BrU stacks more strongly than T in DNA-like conformations, and indeed, in the gas phase, BrU is probably a *weaker* stacker with purines than T is. In water, the differences are simply negligible. We therefore cannot support the hypothesis that the mutagenicity of BrU is due to enhanced stacking.

Note that purines stack more strongly than pyrimidines: the BrU/G dimer optimised from PDB entry 1D9R has the greatest gas-phase potential energy of interaction of any dimer in this study (−17.0 kcal/mol).

Also note that the Gibbs free energies of Py/Pu stacking are mostly unfavourable, except for one gas-phase dimer of thymine and guanine, for which ΔG^{CP} = −0.2 kcal/mol. The Gibbs free energies in water are unfavourable by a virtually uniform amount for all Py/Pu dimers.

5.6. Conclusions

We have calculated the gas-phase interaction energies of a set of stacked nucleobase dimers from the Protein Data Bank, containing 5-bromouracil and/or thymine. This was to test the hypothesis that

BrU stacks more strongly than T, and in that way escapes removal by DNA repair enzymes, even when it mispairs with G through its major tautomer, which forces a stereochemically incorrect base pair (in contrast to the more widely accepted "rare tautomer" model). Two high levels of theory (df-LCCSD(T) and RI-mPW2PLYP-D) indicated that, on average, BrU does stack more strongly than T in almost all base-pair/sequence combinations, but usually by less than 1.25 kcal/mol. The greatest difference in favour of BrU stacking was found when cytosine was on the 5' side. The use of fixed geometries (not energy minima) precluded the calculation of thermal effects.

We then attempted to locate all possible energy minima of every stacked dimer combination of T, BrU, C and U, in both the gas phase and water (PCM), using a previously developed methodology. All the dimers except gas-phase C/C can adopt several energy-minimum geometries, which are distinguished principally by the rotation of one base about the vertical axis connecting the two base midpoints ("twist"), and more subtly, by variations in the angle between the base planes and the extent of their horizontal displacement. The vertical distances between the bases in the optimised geometries were consistently shorter than the known average in B-DNA by around 0.3 Å.

In the gas phase, some structure of C/BrU, T/BrU or U/BrU with a greater binding potential energy than C/T, T/T or U/T, respectively, was always found. Due to their twist angles, many of these dimers differed considerably from anything encountered in DNA; but when only the dimers with the most DNA-like twist were considered, C/BrU and T/BrU were still more strongly bound than C/T and T/T.

However, stacking in the helix is an enthalpy-driven process, and one in which water molecules bound to DNA play a key role. When enthalpic and solvent contributions were taken into account – either together or separately – the stacking advantage of BrU disappeared.

The enthalpy term is not exact: the harmonic approximation introduces an error, particularly in the low-frequency vibrations. Likewise, PCM is at best a rough guide to the effect of water *in vivo*, since it neglects explicit base–water interactions, and assumes a uniform distribution of (bulk) water. Follow-up studies which include explicit solvent molecules at experimental (PDB) locations may therefore be useful. Nonetheless, the general tendency of thermal and solvent effects to weaken the stacking of BrU is clear enough.

The incorporation of BrU into DNA increases the mutation rate by a factor of around 100–10,000.[53, 57, 62, 154, 155] Considering our results, the only slightly better stacking ability of BrU than thymine – which is in any case largely nullified by thermal and aqueous effects – does not seem a plausible explanation for the considerably greater ability of BrU-G mispairs to escape enzymatic repair than T-G mispairs. That is an explanatory power which the wobble pair model of BrU-mutagenesis would require stacking to have, so the wobble pair model is not supported by our findings.

A caveat to this conclusion is that most of the experimental stacked dimers containing BrU came from structures in which BrU was correctly paired with adenine in the opposite strand; only three PDB codes (1DA1, 1UNM and 2OBZ) contained BrU-G mispairs, which were wobble pairs in each case. It is possible that BrU stacks more strongly when wobble-paired with guanine than when correctly paired with adenine, and this would support the wobble pair model of BrU-mutagenesis if the corresponding effect for thymine were weak or absent. However, we could not find enough experimental BrU-G wobble pairs to make such a comparison.

Chapter 6: Conclusions

We have used various methodologies to investigate the keto-enol tautomerism (in water) and the base stacking strength (in water and the gas phase) of the mutagenic DNA base analogue 5-bromouracil. This has been a broad project, in which several lines of enquiry have been followed to establish quantitative grounds for choosing between two rival models of BrU-G mispairing: the conventional "rare tautomer" hypothesis and the wobble pair model. The actual DNA replication process is too complex to be simulated dynamically, and in full quantum detail, with current technology, so it is not possible to simply "insert" BrU into a realistic cellular environment and follow the subsequent mispairing. However, theory can provide powerful circumstantial evidence, *via* the inherent tautomeric, thermodynamic and stereochemical differences between BrU and the natural bases thymine and uracil. This evidence can be used to assess the plausibility of various assumptions which underlie the rival explanations of BrU's propensity to mispair and/or escape removal after mispairing.

It has long been supposed that aqueous solvation affects the tautomerism of BrU differently from the natural bases, favouring the "mutagenic" O4-enol tautomer of BrU. However, the experimental evidence – provided as long ago as 1962 by Katritzky and Waring[65] – has been treated sceptically in the subsequent literature.[30, 215] To put this on firmer theoretical ground, the group of van Mourik, in collaboration with Danilov and Gaigeot, has carried out both static and dynamic calculations on the keto and enol tautomers of BrU solvated by 49–100 explicit water molecules.[87-89] In extending the static studies by carrying out the geometry optimisations at higher levels of theory (Chapter 3), we have shown that aqueous solvation increases the "rare tautomer" population of BrU more than that of U, though perhaps to a lesser extent than originally[87, 88] suggested. Since water is present not only at the surface of DNA itself, but also in the environment from which the building blocks of DNA (substrate nucleosides) are recruited during replication, this suggests that solvation may aid the formation of BrU*-G mispairs *in vivo*.

However, the calculated difference in tautomerisation energies between BrU and U varies widely by method: in 50-water clusters, the difference in favour of BrU tautomerisation (compared to U) was 17.2 kcal/mol with B3LYP in van Mourik's study,[88] but only 0.6 kcal/mol with M05-2X in this thesis. (In 100-water clusters, van Mourik *et al.* even obtained a difference of 23.7 kcal/mol with B3LYP.) This span is too great to allow meaningful quantitative predictions of BrU's behaviour in the cell. Further investigation of *how* water stabilises non-canonical tautomers of BrU, at the quantum level, is also needed, perhaps along the lines of the NBO analysis performed by Lukmanov *et al.* on various other substituted uracils.[142] Electron-withdrawing effects may play a role.

In Chapter 4, molecular dynamics simulations showed that deprotonation of the diketo tautomer at N3 (the first step in intramolecular proton exchange, i.e. tautomerisation) had a more favourable Helmholtz free energy for BrU than for U under aqueous conditions, thus further supporting the rare tautomer hypothesis. The corresponding pK_a of BrU is a valuable result in itself, since it is difficult to measure experimentally. However, further simulations are needed to improve the accuracy of this finding.

In Chapter 5, it was shown that the major tautomer of BrU can undergo somewhat stronger base stacking interactions than T or U in the gas phase at 0 K, but that solvation, thermal effects, and the

geometric requirements of the DNA helix tend to nullify this difference. It was therefore concluded that enhanced base stacking is probably not responsible for the durability of BrU-G mispairs, contrary to that suggestion in support of the wobble pair (i.e. major-tautomeric) model. However, there was considerable variation in the relative stacking strengths of BrU and T, depending on the dimer geometry and the identity of the stacking partner.

The finding that BrU stacks more strongly than T by around 2 kcal/mol when positioned 3' to a cytosine residue (albeit only in terms of gas-phase potential energy) suggests a possible sequence specificity in BrU-induced mutations. A template BrU residue appearing in the sequence 5'-C/BrU-3' should be especially mutable, since the strong stacking of this sequence would stabilise any mispairs that appeared due to misinsertion of G in the nascent strand. Conversely, if BrU were a substrate nucleoside, then the 5' G in any template GG sequence should be mutable, because misinsertion of BrU opposite it would create highly stacked 5'-C/BrU-3' in the nascent strand. Intriguingly, Davidson *et al.* did observe that 95% of the GC→AT transitions induced by substrate BrU in mouse cells occurred in GG sequences.[63] However, in 90% of those cases, the mutated G was on the 3' side, which corresponds to the sequence 5'-BrU/C-3' in the nascent strand. In contrast to 5'-C/BrU-3', our mean stacking energies of 5'-BrU/C-3' in experimental geometries were almost identical to those of 5'-T/C-3', according to both df-LCCSD(T) and RI-mPW2PLYP-D (Chapter 5, Tables 5.3 and 5.4). The high mutability of Davidson *et al.*'s 3' G residues therefore cannot be explained by our stacking results. What would have been a fascinating concordance between the observed sequence specificity of mutation and our calculated stacking energies was unfortunately not to be.

We note in passing that the stacking explanation of BrU mutagenicity, even if verified, could not necessarily be applied to other mutagenic agents. For example, the adenine analogue 2-aminopurine causes transition mutations by mispairing with cytosine, but it seems unlikely that the changed position of the amino group (from adenine, which is 6-aminopurine) could so greatly enhance the stacking of this compound.

In this thesis, we have not investigated the ionisation model of BrU-G mispairing. It is closely related to the rare tautomer model, since it invokes the same stereochemistry (pseudo-Watson–Crick; see Chapter 1, Figure 1.9), and depends on deprotonation of BrU at N3, to allow this site to accept a hydrogen bond from N1-H1 of guanine. The first targeted investigation of BrU-G mispairing treated the two models as indistinguishable,[79] and indeed it is hard to discriminate between them using structural evidence alone, since hydrogen atoms are absent from x-ray data. The lower pK_a of BrU than U, for which evidence was adduced in Chapter 4, is consistent with BrU having a greater stability of its ionised form than uracil does, and could be interpreted as supporting both models equally. However, it has been argued that the unfavourable interaction between the δ-negative O4(BrU$^-$) and O6(G) atoms prevents the incorporation of a planar ionised BrU-G mispair into DNA.[70, 216] We cannot comment further based on the results in this thesis.

To ask why BrU mispairs more often than T is to ask why it has a lower specificity for adenine. Consider that in the early days of life, Nature selected A, C, G, T and U as the carriers of genetic information – presumably in preference to other contenders – due to their extremely high fidelity of base pairing. It is therefore not surprising that a different molecule, which simply happens to bear a steric resemblance to thymine, should be less exquisitely adapted for correct pairing – but this could be for more than one "chemical" reason. It is quite possible that BrU mispairs with G in multiple ways – perhaps all of the candidate structures described in Chapter 1 occasionally play a role in mutation, and their relative frequency depends on the local sequence environment and on biochemical

conditions within the cell. We note the findings of Sowers *et al.*, that the mispairs of 5-bromouracil, 5-fluorouracil and 5-chlorouracil with guanine, in short synthetic stretches of DNA, all exhibited pH-dependent equilibria between wobble and ionised structures (though they were not subsequently involved in mutagenesis).[83, 217, 218]

The results herein provide overall support for the rare tautomer model, but the complicated set of experimental observations about BrU mutagenicity (its variable rate, sequence dependence, and the ratio of GC→AT to AT→GC transitions) may not be explicable by a single chemical datum, such as its tautomeric ratio. For example, suppose that further investigations put it beyond doubt that the rare tautomer population of BrU, under cellular conditions, was higher than that of U, but only by a factor of 100 (corresponding to a tautomerisation energy difference of roughly 3 kcal/mol between the two bases, cf. Equation 1.1).[19, 154] This alone would not be sufficient to explain the mutagenicity of BrU, nor its sequence dependence, at least according to any straightforward reasoning. (2 kcal/mol is the upper bound of the ability of *stacking* to stabilise BrU-G mispairs, being roughly the difference between the gas-phase interaction energies of 5'-C/BrU-3' and 5'-C/T-3' in experimental geometries.)

Throughout this thesis, we have assumed that BrU causes mutations by mispairing with G, and have focused on the intrinsic chemical differences between BrU, T and U, to elucidate why BrU-G is especially likely to form, and in what stereochemistry. However, it is possible that mechanisms which do not exclusively involve the mispairing of BrU also play a role in mutagenesis. We have seen that Kaufman and Davidson proposed deoxycytidine starvation in the nucleotide pool as a way for BrU to promote misinsertion of thymine opposite guanine, without itself being incorporated into DNA.[60] Interestingly, Vargason *et al.* suggested that bromination of cytosine at the 5-position could increase the mutation rate by promoting an unusual helical geometry ("E-DNA"), which facilitates the water-assisted deamination of cytosine to uracil, causing GC→AT transitions;[219] perhaps bromination of uracil could also have destabilising structural effects. BrU mutagenesis is often treated as a special case of the general phenomenon of spontaneous point mutations, which are assumed to arise from G-T and A-C mispairs – although even for the natural bases, the structures of the mispairs remain unclear, many decades after Watson and Crick's suggestion of rare tautomers. However, it is not self-evident that the mutagenic action of BrU must be simply an accelerated version of the mechanism involving thymine. It is to be expected that any unnatural base analogue could interfere with the smooth functioning of DNA, simply by its unwanted presence in the cellular machinery – but this could be a chaotic effect, involving perturbations to more than one step in the replication process.

A hint at the complexity of the mutagenic activity of 5-bromouracil is offered by the closely related compound 5-fluorouracil (FU), an anti-cancer drug. One of its cytotoxic mechanisms is inhibitory binding[220] of the enzyme thymidylate synthase, causing a fatal lack of thymidine residues available during cell division (cf. the cytidineless state[60] apparently induced by BrU). But FU can also be incorporated directly into both RNA and DNA, where it interferes with numerous cellular processes – whether this involves mispairing with guanine remains unclear.[72, 82, 220] To examine, in all-atom detail, processes such as the enzymatic proofreading of BrU-G mismatches, or the binding of synthase enzymes by FU, should be the next challenge in extending the scope of the methodologies used in this thesis.

Appendix A: Proof of the first Hohenberg–Kohn theorem by *reductio ad absurdum*

Assume that two different potentials, ν and ν′, arising from two different nuclear geometries, give rise to the same electron density, ρ.

Now, the potentials determine the corresponding Hamiltonians and wavefunctions:

$$\nu \rightarrow \hat{\mathrm{H}} \rightarrow \Psi;\ \nu' \rightarrow \hat{\mathrm{H}}' \rightarrow \Psi' \tag{A.1}$$

By the variational principle, E_0 – the ground-state energy of the system with potential ν – is lower than the energy obtained by the operation of $\hat{\mathrm{H}}$ on the wavefunction of the system with potential ν′:

$$E_0 = \langle \Psi | \hat{\mathrm{H}} | \Psi \rangle < \langle \Psi' | \hat{\mathrm{H}} | \Psi' \rangle = \langle \Psi' | \hat{\mathrm{H}}' | \Psi' \rangle - \langle \Psi' | \hat{\mathrm{H}} - \hat{\mathrm{H}}' | \Psi' \rangle \tag{A.2}$$

where $\langle \Psi' | \hat{\mathrm{H}}' | \Psi' \rangle = E_0'$ is the ground-state energy of the system with potential ν′, and the quantity $\langle \Psi' | \hat{\mathrm{H}} - \hat{\mathrm{H}}' | \Psi' \rangle$ is the difference between the effect of the two Hamiltonians operating on Ψ′. Therefore:

$$E_0 < E_0' - \int \rho(\boldsymbol{r}) \left[\nu(\boldsymbol{r}) - \nu'(\boldsymbol{r}) \right] \mathrm{d}\boldsymbol{r} \tag{A.3}$$

Starting from E_0', and proceeding in the same way, we can derive:

$$E_0' < E_0 - \int \rho(\boldsymbol{r}) \left[\nu'(\boldsymbol{r}) - \nu(\boldsymbol{r}) \right] \mathrm{d}\boldsymbol{r} \tag{A.4}$$

Now, since

$$\int \rho(\boldsymbol{r}) \left[\nu(\boldsymbol{r}) - \nu'(\boldsymbol{r}) \right] \mathrm{d}\boldsymbol{r} + \int \rho(\boldsymbol{r}) \left[\nu'(\boldsymbol{r}) - \nu(\boldsymbol{r}) \right] \mathrm{d}\boldsymbol{r} = 0 \tag{A.5}$$

it follows by addition of the two inequalities that

$$E_0 + E_0' < E_0' + E_0 \tag{A.6}$$

which is logically contradictory. Therefore, the assumption that two different potentials can give rise to the same density is impossible, hence the potential uniquely determines the density.

Appendix B: List of PDB entries used in Chapter 5

Tables B.1 and B.2: PDB entries used to generate the experimental stacked structures in Chapter 5, Section 5.2 (published as Supporting Information to L. Holroyd, T. van Mourik, *Theor. Chem. Acc.* **133** (2014) 1431).

5-Bromouracil stacking with adenine, cytosine, guanine and thymine:

115D-BrUA	1CS7-GBrU	1FN1-GBrU	1SM5-BrUT	2H05-GBrU	2W7N-CBrU
115D-GBrU	1D3R-CBrU	1GQU-ABrU	1SM5-TBrU	2OBZ-GBrU	344D-GBrU
1A35-ABrU	1D9R-ABrU	1JXL-BrUC	1UNM-BrUT	2OBZ-BrUC	366D-BrUA
1A35-BrUC	1D9R-BrUG	1JXL-TBrU	1WTP-ABrU	2W36-BrUA	366D-GBrU
1A35-BrUT	1DA1-BrUG	1K8P-BrUA	1WTP-BrUC	2W36-CBrU	3GDD-BrUA
1A35-TBrU	1DCR-ABrU	1P1Y-ABrU	2DLJ-BrUA	2W7N-ABrU	3OH9-ABrU
1AIO-CBrU	1DCR-BrUG	1P1Y-BrUG	2DLJ-GBrU	2W7N-BrUG	3OH9-BrUG
1CS7-BrUT	1FN1-BrUA	1P54-BrUG	2H05-BrUA	2W7N-BrUT	

Table B.1

Thymine stacking with adenine, cytosine, guanine and thymine:

1A35-AT	1CS7-TG	2L5K-GT	2W7N-AT	3OH9-TC	3PW5-TC
1A35-CT	1CS7-TT	2L5K-TC	2W7N-CT	3OH9-TG	3PW5-TT
1A35-GT	1VTJ-AT	2L5K-TT	2W7N-TG	3OH9-TT	3R86-GT
1A35-TT	1VTJ-TC	2L8P-AT	2W7N-TT	3OIE-AT	3R86-TA
1AIO-CT	2L2V-CT	2L8P-TG	3G6X-GT	3OIE-TC	3U89-CT
1AIO-TG	2L2V-TA	2LGM-GT	3G6X-TC	3OIE-TT	3U89-TG
1BNA-AT	2L2V-TG	2LGM-TA	3IXN-GT	3OPI-TC	
1BNA-TC	2L2V-TT	2ORF-GT	3IXN-TA	3OPI-TT	
1BNA-TT	2L5K-AT	2ORF-TA	3OH9-AT	3PW5-AT	

Table B.2

Format: AAAA-XY where AAAA is the four-character PDB code, X is the base nearer the 5' end of the strand and Y is the base nearer the 3' end.

Bases: A = adenine, BrU = 5-bromouracil, C = cytosine, G = guanine, T = thymine.

Note that in many PDB entries, a given sequence XY occurs more than once. We have only used the first instance of any given sequence found in each PDB entry.

Appendix C: Details of scans in Chapter 5

C.1. BrU/BrU

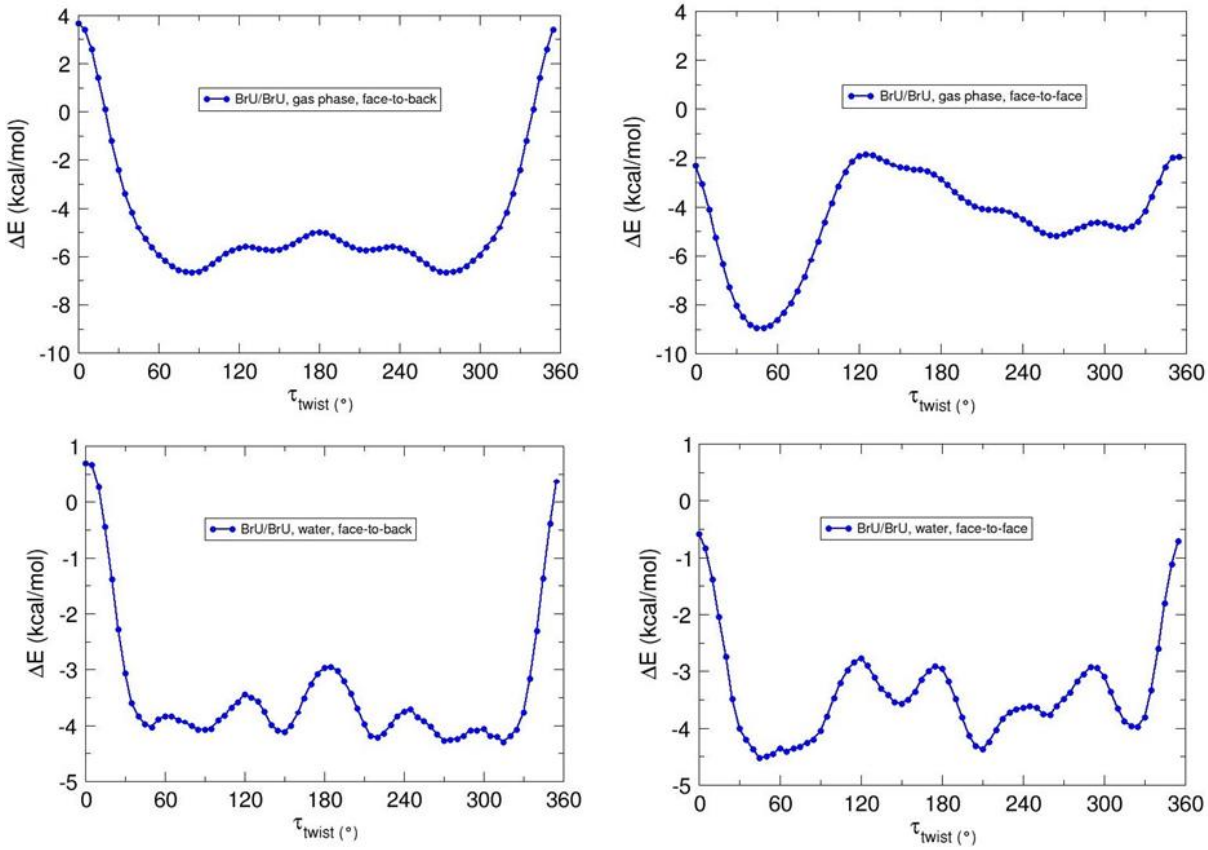


Figure C.1: Interaction energies (Δ*E*) of BrU/BrU as a function of the twist angle τ.

Figure C.1 (top left) shows the face-to-back (FTB) gas-phase interaction energy of BrU/BrU as a function of rigidly scanning the "twist" angle τ (Chapter 5, Section 5.3). Due to the identical monomers, the rotational profile is symmetric about τ = 180°. There are four minima, at τ = 85°, 145°, 215° and 275°, and four maxima, at τ = 0°, 125°, 180° and 235°. The BSSE (not shown) ranges from 5.58 to 6.79 kcal/mol: almost double that of U/BrU. The interaction energy at the highest maximum (0°) is strongly repulsive: here, ΔE^{CP} = +3.7 kcal/mol (but this value would be negative if the counterpoise correction were neglected). At this geometry, all the substituents on one base (including Br and C=O) are situated exactly above the corresponding substituents on the other base, at a distance of 3.4 Å.

Full optimisation of the gas-phase structures yielded just three energy minima, because the optimisations starting from τ = 145° and τ = 215° both converged to the same structure, with τ = 180° and ΔE_0^{CP} = −10.20 kcal/mol. The other two minima form a symmetry-equivalent pair, with optimised τ angles of 73° and 287°, and ΔE_0^{CP} = −9.50 kcal/mol for both. Three transition states were also found, of which two form a symmetry-equivalent pair.

Figure C.1 (bottom left) shows the face-to-back interaction energy of BrU/BrU in water PCM solvent. As in the gas phase, the profile is symmetric. However, it is flatter (compare the two *y*-axes). There are six minima, at 50°, 90°, 150°, 220°, 270° and 315°, and six maxima, at 0°, 65°, 120°, 185°, 245° and 300°.

In water, the optimisations yielded six structures, falling into three symmetry-equivalent pairs. The ZPE-corrected interaction energies of all six optimised dimers fall within the narrow range of −5.0 to −5.4 kcal/mol, the pair with τ = 60°/300° having the greatest value. Transition state searches between adjacent optimised minima all successfully yielded structures with one imaginary frequency. However, the TS found by using the minima with τ = 243° and τ = 260° had an optimised τ value of 273°, and was identical to the TS found by using the minima with τ = 260° and τ = 300°. "Following the frequency" of this TS yielded the latter pair of minima, confirming that it links those two structures, rather than the former pair.

Figure C.1 (top right) shows the face-to-face (FTF) gas-phase stacking of BrU/BrU. Unlike in the FTB case, there is no symmetry. The three minima are at 50°, 265° and 315°, the three maxima at 125°, 295° and 355°. There are no regions of repulsive interaction. The BSSE ranges from 5.50 to 6.82 kcal/mol.

Figure C.1 (bottom right) shows the FTF aqueous stacking of BrU/BrU. Excluding a very shallow minimum at 65°, there are five minima, at 45°, 150°, 210°, 260° and 325°. The corresponding maxima are at 0°, 120°, 175°, 245° and 290°. Again there are no repulsive regions.

Optimisation of the gas-phase structures yielded only two minima, as the starting structures with τ = 265° and 315° converged to the same minimum, with τ = 295° and ΔE_0^{CP} = −8.3 kcal/mol. The other minimum, with τ = 49° and ΔE_0^{CP} = −11.8 kcal/mol, has the greatest interaction energy of any gas-phase BrU/BrU dimer found in this study. Two transition states were located, but the one found by starting from τ = 125° was L-shaped rather than stacked, and probably does not connect the two minima.

Optimisation of the aqueous structures initially yielded five distinct minima, of which the most strongly bound had τ = 56° and ΔE_0^{CP} = −5.6 kcal/mol. However, the minimum with τ = 121° (obtained from the starting structure with τ = 150°) could only be found by using Gaussian's default convergence criteria – when "Tight" criteria were used, only structures with one imaginary frequency were found when starting a minimum search from τ = 150°. Subsequently, five transition states were located in the usual way. However, the guess structures with τ = 120° and 175° both yielded the same TS, with optimised τ = 120°. Furthermore, this TS was lower in energy (ΔE_0^{CP} = −5.0 kcal/mol) than the minimum with τ = 121° (ΔE_0^{CP} = −4.8 kcal/mol), which was one of the minima it was expected to connect. Nevertheless, "following the frequency" of this TS with displacements in one direction yielded the minimum with τ = 121° (using "Tight" optimisation criteria). With displacements in the other direction, a sixth energy minimum was found, which, despite its relatively large angle between planes (12°) and horizontal separation (2.66 Å) had a stronger interaction energy than the other five structures (ΔE_0^{CP} = −5.9 kcal/mol).

C.2. C/BrU

Figure C.2 (top left) shows the gas-phase FTB stacking of C/BrU. There are two minima, at 40° and 260°, of which the second is much deeper. The two maxima, at 0° and 120°, have similar heights. There are no repulsive regions. The BSSE varies from 3.09 to 3.85 kcal/mol.

Figure C.2 (bottom left) shows the aqueous FTB stacking of C/BrU. There are five minima, at 40°, 85°, 155° 210° and ~285°, the last of which is very wide and flat and is also the deepest. The maxima are at 0°, 60°, 120°, 175° and 240°.

Full optimisation of the gas phase structures yielded two minima (with τ = 52° and 261°) and two transition states. As in the rotational scan, the latter of the two minima is considerably deeper, with ΔE_0^{CP} = −14.0 kcal/mol for the structure with τ = 261°, compared to −7.6 kcal/mol for the structure with τ = 52°.

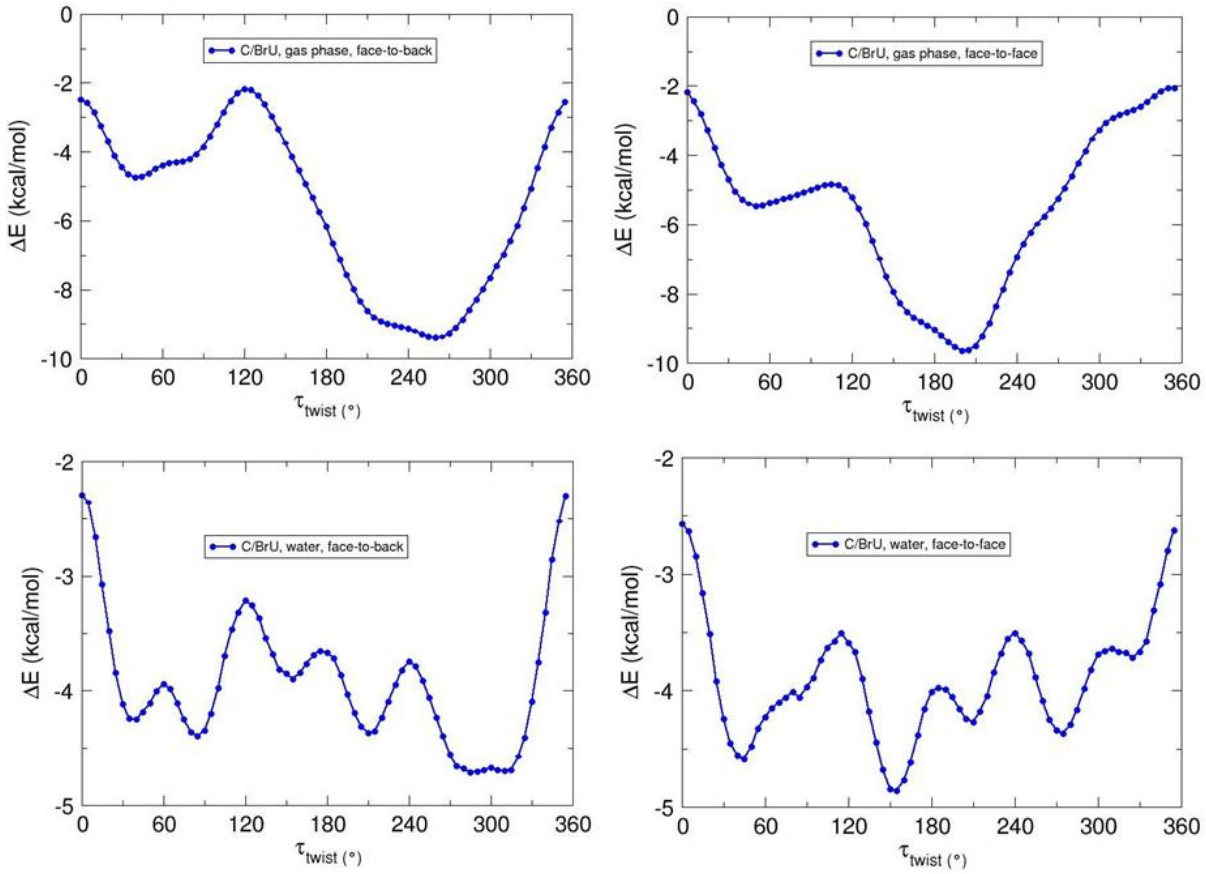


Figure C.2: Interaction energies (Δ*E*) of C/BrU as a function of the twist angle τ.

Full optimisation of the aqueous structures initially yielded only four minima, as the optimisations starting from τ = 40° and 85° converged to the same structure, with τ = 54°. Hence only four transition states could be located. However, the transition states with τ = 329° and 79°, which would each be expected to connect the minimum with τ = 54° to another minimum, were lower in energy than the τ = 54° structure. "Following the imaginary frequency" of these two transition states yielded two new minima each, so that eight distinct stacked minima were found in total. Two of these, with τ = 67° and 293°, are symmetry-equivalent, and their interaction energies (ΔE_0^{CP} = −6.4 kcal/mol for both) are the greatest among these structures.

Figure C.2 (top right) shows the FTF gas-phase rotational profile of C/BrU. There are two minima, at 50° and 200°, with the maxima at 105° and 355°. There are no repulsive regions.

Figure C.2 (bottom right) shows the FTF aqueous rotational profile of C/BrU. There are five minima, at 45°, 155°, 210°, 275° and 325°, with the maxima at 0°, 115°, 185°, 240°and 310°. There are no repulsive regions.

Full optimisation of the gas-phase structures yielded two minima, with τ = 63° and 201°, and two transition states. The minimum with τ = 201° is the more strongly bound.

Full optimisation of the aqueous structures initially yielded five minima and five transition states. However, the TS with $\tau = 163°$ was essentially isoenergetic with the minimum with $\tau = 157°$ and lower in energy than the minimum with 214°, while the TS with $\tau = 274°$ was essentially isoenergetic with the minimum with $\tau = 214°$ and lower in energy than the minimum with $\tau = 300°$. In each case, "following" the imaginary frequency of the TS yielded one new minimum and one that had been located already. Hence seven distinct minima were found in total, of which the most strongly bound has $\tau = 55°$.

C.3. T/BrU

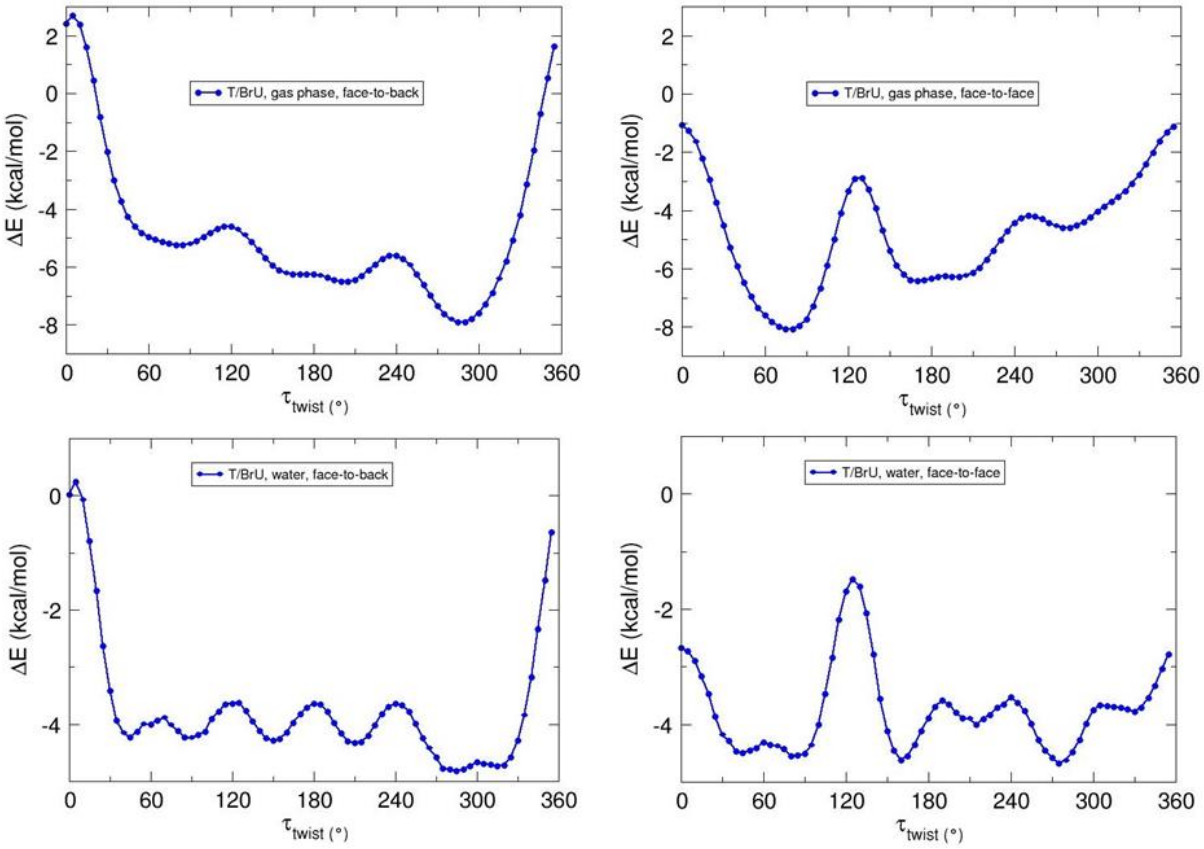


Figure C.3: Interaction energies (Δ*E*) of T/BrU as a function of the twist angle τ.

Figure C.3 (top left) shows the FTB gas-phase rotational profile of T/BrU. There are three minima, at 85°, 205° and 285°, with maxima at 5°, 115° and 240°. The region around the maximum at 5° is repulsive. The BSSE ranges from 3.33 to 3.93 kcal/mol.

Figure C.3 (bottom left) shows the FTB aqueous rotational profile of T/BrU. The six minima are at 45°, 90°, 150°, 210°, 285° and 315°, with the maxima at 5°, 70°, 125°, 180°, 240° and 305°. As in the gas phase, the region around the maximum at 5° is repulsive.

Full optimisation of the gas-phase structures yielded three minima and three transition states, the most strongly interacting dimer being the minimum with $\tau = 289°$, for which $\Delta E_0^{CP} = -11.1$ kcal/mol.

Full optimisation of the aqueous structures yielded only four minima, as the starting structures with $\tau = 90°$ and 150° both converged to a minimum with $\tau = 112°$, while those with $\tau = 285°$ and 315° both found a minimum with $\tau = 291°$. Therefore only four transition states were located. The one with $\tau =$

263° was lower in energy than the minimum with τ = 223°, but "following" the frequency of this TS did not yield any new minima. The global minimum has τ = 291° and ΔE_0^{CP} = −5.9 kcal/mol.

Figure C.3 (top right) shows the FTF gas-phase rotational profile of T/BrU. There are three minima, at 80°, ~170° and 275°, the middle one being very wide and flat. The interaction energies are attractive throughout.

Figure C.3 (bottom right) shows the FTF aqueous rotational profile of T/BrU. There are six minima, at 45°, 80°, 160°, 215°, 275° and 330°, with maxima at 0°, 60°, 125°, 190°, 240° and 305°. The interaction energies are attractive throughout.

Full optimisation of the gas-phase structures yielded three minima. However, the minimum obtained by starting with τ = 275° was not stacked but L-shaped, with an N1H1(Br)•••O2(T) hydrogen bond. The most strongly bound structure was the stacked minimum with τ = 76° and ΔE_0^{CP} = −11.2 kcal/mol.

Full optimisation of the structures in water initially yielded five minima, since the starting structures with τ = 45° and 80° both converged to a structure with τ = 57°. Therefore five transition states were located. However, the TS with τ = 274° had a lower ZPE-corrected energy than the minimum with τ = 225; "following" its imaginary frequency yielded a sixth stacked minimum with τ = 300°. The most stable structure has τ = 57° and ΔE_0^{CP} = −5.7 kcal/mol.

C.4. C/C

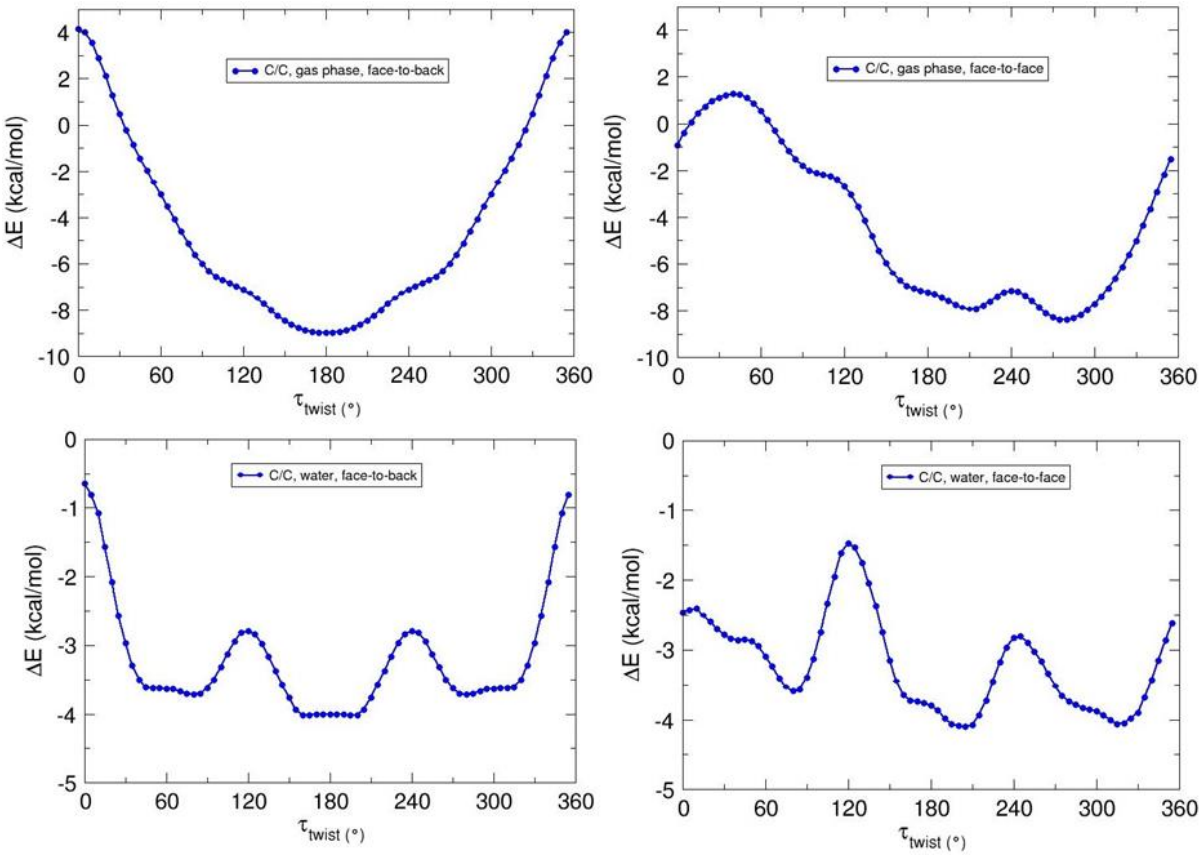


Figure C.4: Interaction energies (Δ*E*) of C/C as a function of the twist angle τ.

Figure C.4 (top left) shows the FTB gas-phase energy profile of the C/C stacked dimer. In contrast to all other dimers explored in this study, there is only one energy minimum, at $\tau = 180°$. The profile is symmetric about this minimum. The maximum is at $\tau = 0°$. A large part of the energy profile around the maximum is positive (repulsive). The BSSE is relatively small and geometry-insensitive, varying from 1.01 to 1.07 kcal/mol.

Figure C.4 (bottom left) shows the FTB aqueous energy profile of the C/C stacked dimer. There are three minima, at 80°, ~180° and 280°, the second of which is especially broad and flat. The maxima are at 0°, 120° and 240°. The interaction energies are negative throughout.

Full optimisation of the gas-phase minimum yielded a significantly tilted structure (angle between planes = 17°) with $\Delta E_0^{CP} = -11.1$ kcal/mol.

For the aqueous structures, we decided to start optimisations from both of the very shallow sub-minima at 160° and 200° within the broad minimum centred on 180°. These optimisations converged to distinct but symmetry-equivalent, heavily tilted structures with $\tau = 141°$ and 219°, angle between planes = 26° and $\Delta E_0^{CP} = -4.1$ kcal/mol. The starting structures with $\tau = 80°$ and 280° converged to symmetry-equivalent, also heavily tilted minima with $\tau = 99°$ and 261°, angle between planes = 17° and $\Delta E_0^{CP} = -4.4$ kcal/mol. Four transition states were located.

Figure C.4 (top right) shows the FTF gas-phase energy profile of the C/C stacked dimer. There are two minima, at 210° and 280°, and two maxima, at 40° and 240°. The region around the first maximum is repulsive.

Figure C.4 (bottom right) shows the FTF aqueous energy profile of the C/C stacked dimer. There are four minima, at 40°, 80°, 205° and 315°, the first of which is very shallow. The maxima are at 10°, 45°, 120° and 245°. The interaction energies are negative throughout.

Full optimisation of the gas-phase structure with $\tau = 210°$ yielded a stacked energy minimum with $\tau = 152°$ and ΔE_0^{CP} −12.6 kcal/mol. However, full optimisation of the structure with $\tau = 280°$ yielded a planar, hydrogen-bonded (non-Watson–Crick) minimum with $\Delta E_0^{CP} = -20.9$ kcal/mol. We attempted to locate a stacked minimum corresponding to $\tau = 280°$ by repeating the geometry optimisations with MP2/6-31+G(d) and B97-D/6-31+G(d), but these procedures followed the same path as M06-2X. Following these findings, no transition state searches were attempted.

Full optimisation of the aqueous structures initially yielded only three minima, because the attempt to optimise the structure with $\tau = 80°$ to a minimum in fact converged to a structure with one imaginary (negative) vibrational frequency. However, "following" the motion of this vibration yielded a true minimum, as confirmed by the absence of negative vibrational frequencies in the resulting structure, which has a τ angle of 102°. The structure with the greatest ZPE-corrected interaction energy has $\tau = 310°$ and $\Delta E_0^{CP} = -4.4$ kcal/mol.

C.5. C/T

Figure C.5 (top left) shows the FTB gas-phase energy profile of the C/T dimer. There are two minima, at 85° and 210°, and two maxima, at 5° and 115°. The interaction energies are negative throughout. The BSSE is relatively small and geometry-invariant, ranging from 1.00 to 1.09 kcal/mol.

Figure C.5 (bottom left) shows the FTB aqueous energy profile of the C/T dimer. There are five minima, at 80°, 160°, 205°, 285° and 315°, with the maxima at 0°, 120°, 170°, 240° and 290°. The interaction energies are negative throughout.

Full optimisation of the gas-phase dimers yielded two minima, the more strongly bound of which had $\tau = 222°$ and $\Delta E_0^{CP} = -10.9$ kcal/mol. Two transition states were also located. However, both were lower in energy than the minimum with $\tau = 84°$, and one of them (with $\tau = 293°$) was L-shaped. "Following" the imaginary frequency of the TS with $\tau = 149°$ yielded two stacked minima: one was new (with $\tau = 132°$), the other was identical to the already-found structure with $\tau = 222°$. Carrying out the same procedure for the L-shaped TS with $\tau = 293°$ yielded two unstacked minima: one also L-shaped, containing an N4H4(C)•••O4(T) hydrogen bond, the other planar, with the bases bound by two hydrogen bonds in non-Watson–Crick fashion.

Full optimisation of the aqueous dimers initially yielded four minima, since the starting structures with $\tau = 285°$ and 315° converged to the same minimum, with $\tau = 296°$. Therefore four transition state searches were launched. However, one of these (starting from the guess state with $\tau = 120°$) returned a new energy minimum. Also, "following" the imaginary frequency of the TS with $\tau = 230°$ returned two new energy minima. Hence seven minima were found in total, of which the most strongly bound has $\tau = 296°$ and $\Delta E_0^{CP} = -5.8$ kcal/mol.

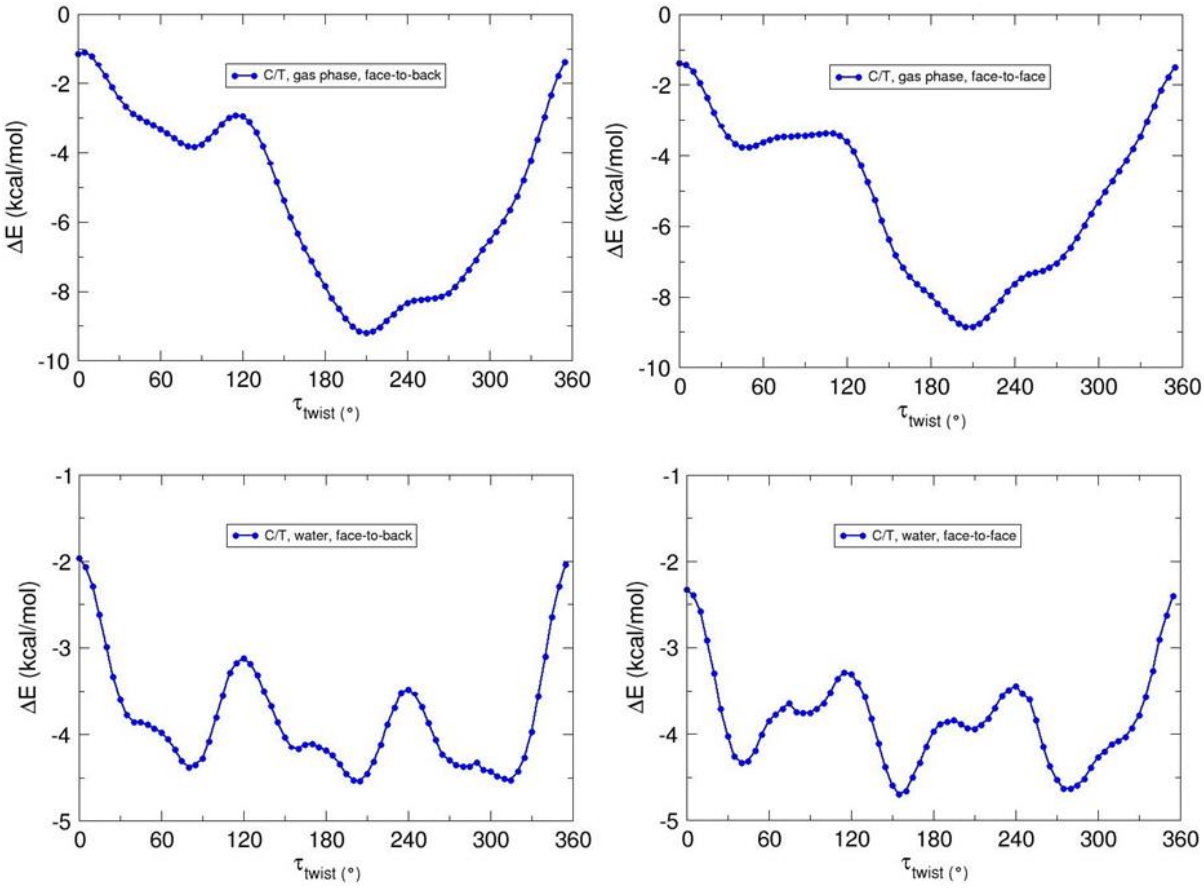


Figure C.5: Interaction energies (Δ*E*) of C/T as a function of the twist angle τ.

Figure C.5 (top right) shows the FTF gas-phase rotational profile of stacked C/T. There are two minima, at 50° and 210°. The maxima are at 0° and 105°. The interaction energies are negative throughout.

Figure C.5 (bottom right) shows the FTF aqueous rotational profile of stacked C/T. There are five minima, at 40°, 85°, 155°, 210° and 275°. The interaction energies are negative throughout.

Full optimisation of the gas-phase structures initially yielded only one minimum (with τ = 209°), because the attempt to locate a minimum starting from τ = 50° failed to find a stationary point when using Gaussian's "Tight" convergence criteria. However, with the default criteria, a structure with one imaginary vibrational frequency was located. Then, "following" the imaginary frequency yielded a true minimum, with τ = 79°. Two transition states were then located in the attempt to connect the two minima. However, the one with τ = 158° was lower in energy than the minimum with τ = 79°. "Following" the imaginary frequency of this structure yielded two new minima: one stacked, the other L-shaped, with an N1H1(C)•••O4(T) hydrogen bond. The most strongly bound structure has τ = 209° and ΔE_0^{CP} = −12.1 kcal/mol.

Full optimisation of the aqueous structures initially yielded five minima and five transition states. However, "following" the imaginary frequency of the TS with τ = 357° yielded two additional minima. The most strongly bound structure in terms of non-ZPE-corrected energy is the minimum with τ = 61°, which has ΔE^{CP} = −6.2 kcal/mol. However, after ZPE corrections, the most strongly bound structure is the TS with τ = 88°, which has ΔE_0^{CP} = −5.3 kcal/mol, compared to 5.2 kcal/mol for the minimum with τ = 61°. This anomalous result is probably due to inaccuracies in the calculation of zero-point corrections, mostly resulting from the use of the harmonic approximation.

C.6. C/U

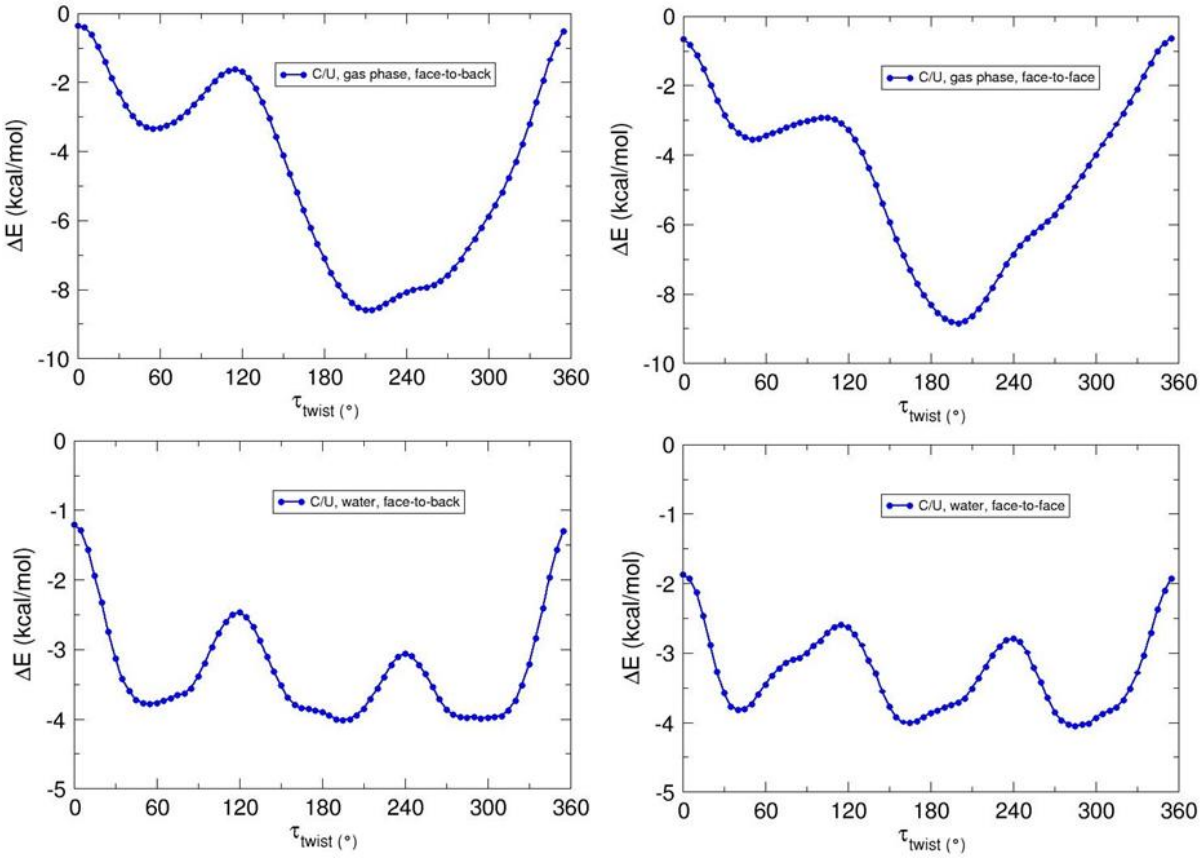


Figure C.6: Interaction energies (Δ*E*) of C/U as a function of the twist angle τ.

Figure C.6 (top left) shows the gas-phase FTB energy profile of stacked C/U. There are two minima, at 55° and 210°, with maxima at 0° and 115°. The interaction energies are negative throughout. The BSSE is fairly small and geometry-invariant, ranging from 0.95 to 1.03 kcal/mol.

Figure C.6 (bottom left) shows the aqueous FTB energy profile of stacked C/U. There are three minima, at 55°, 195° and 295°, with maxima at 0°, 120° and 240°. The interaction energies are negative throughout.

Full optimisation of the gas-phase structures yielded two minima and two transition states. The more strongly bound minimum has $\tau = 220°$ and $\Delta E_0^{CP} = -10.2$ kcal/mol.

Full optimisation of the aqueous structures initially yielded only two minima, because the attempt to optimise from $\tau = 55°$ converged to a structure with one imaginary vibrational frequency. However, "following" this frequency yielded two new structures – one, geometrically very similar, also featuring a single imaginary frequency, the other a true minimum, with $\tau = 60°$. Using the three obtained minima as inputs, three further TSs were located. The most strongly bound minimum has $\tau = 295°$ and $\Delta E_0^{CP} = -5.0$ kcal/mol.

Figure C.6 (top right) shows the gas-phase FTF energy profile of stacked C/U. There are two minima, at 50° and 200°, with maxima at 105° and 355°. The interaction energies are negative throughout.

Figure C.6 (bottom right) shows the gas-phase FTF energy profile of stacked C/U. There are three minima, at 40°, 165° and 285°, with maxima at 0°, 115° and 240°. The interaction energies are negative throughout.

Full optimisation of the gas-phase structures initially yielded two minima and two transition states. However, the TS with $\tau = 160°$ was lower in energy than the minimum with $\tau = 80°$ – "following" the negative frequency of this TS yielded one new stacked minimum and one planar structure, which was hydrogen-bonded in non-Watson–Crick geometry. The most strongly bound stacked structure has $\tau = 209°$ and $\Delta E_0^{CP} = -12.6$ kcal/mol.

Full optimisation of the aqueous structures initially yielded three minima and three transition states. However, "following" the imaginary frequency of the TS with $\tau = 356°$ yielded two additional stacked minima. The most strongly bound minimum has $\tau = 77°$ and $\Delta E_0^{CP} = -4.5$ kcal/mol.

C.7. T/U

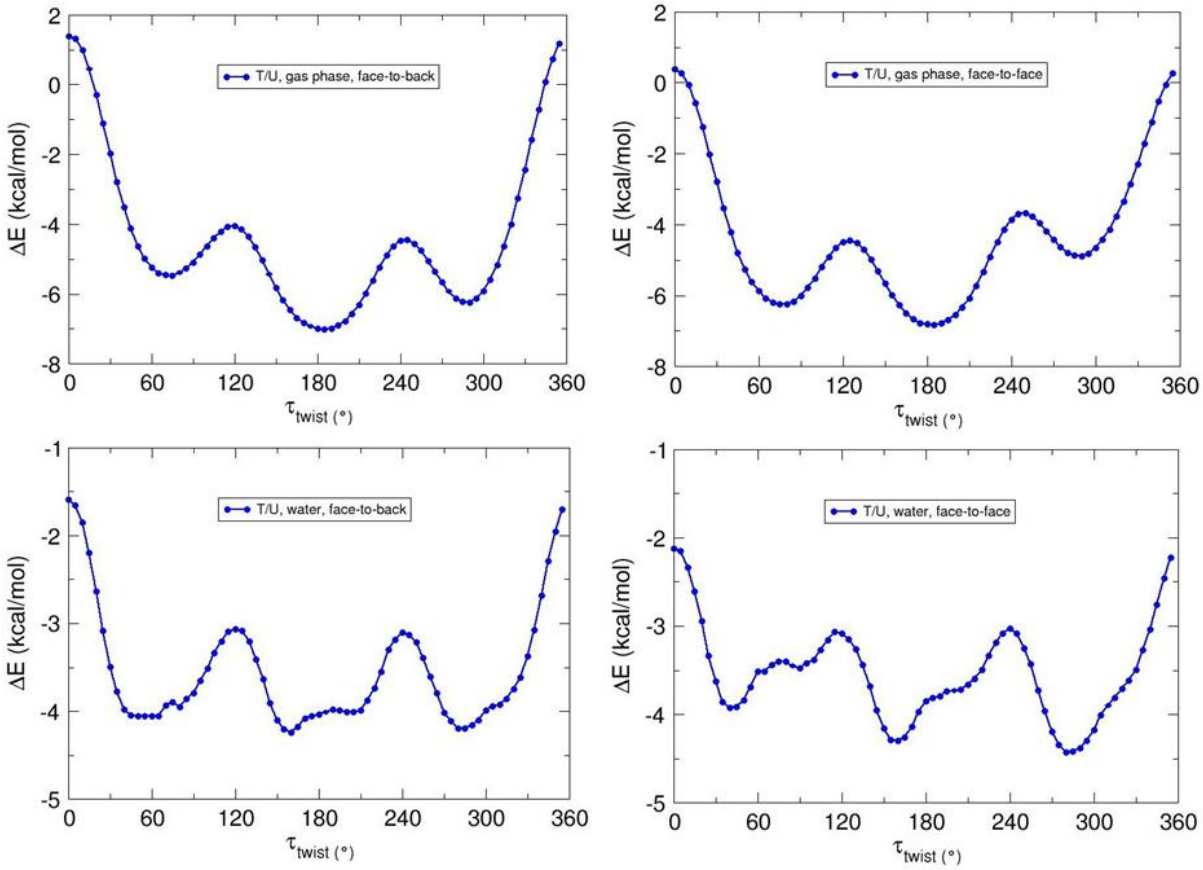


Figure C.7: Interaction energies (Δ*E*) of T/U as a function of the twist angle τ.

Figure C.7 (top left) shows the FTB gas-phase rotational energy profile of stacked T/U. There are three minima, at 75°, 185° and 290°, with the maxima at 0°, 120° and 245°. Interaction energies are positive around the maximum at 0°. The BSSE varies from 0.93 to 1.01 kcal/mol.

Figure C.7 (bottom left) shows the FTF gas-phase rotational energy profile of stacked T/U. There are three minima, all rather broad, at ~60°, ~180° and ~285°, with the maxima at 0°, 120° and 240°. Interaction energies are negative throughout.

Full optimisation of the gas-phase structures yielded three minima and three transition states. The most strongly bound minimum has τ = 182° and ΔE_0^{CP} = −9.1 kcal/mol.

For the aqueous structures, we decided to attempt optimisations from both of the shallow sub-minima, with τ = 160° and 205°, within the broader minimum centred on τ = 180°. Hence four geometry optimisations were launched, yielding four distinct minima. Four transition states were subsequently located. The strongest interaction energy is for the dimer with τ = 66°, which has ΔE_0^{CP} = −4.9 kcal/mol.

Figure C.7 (top right) shows the FTF gas-phase rotational energy profile of stacked T/U. There are three minima, at 75°, 185° and 290°, with the maxima at 0°, 120° and 250°. Interaction energies are slightly positive around the maximum at 0°.

Figure C.7 (bottom right) shows the FTF aqueous rotational energy profile of stacked T/U. There are four minima, at 40°, 90°, 160° and 280°, with the maxima at 0°, 80°, 115° and 240°. Interaction energies are negative throughout.

Full optimisation of the gas-phase structures yielded three minima and three transition states. The most strongly bound minimum has $\tau = 75°$ and $\Delta E_0^{CP} = -9.3$ kcal/mol.

Full optimisation of the aqueous structures yielded four minima and four transition states. The most strongly bound minimum has $\tau = 62°$ and $\Delta E_0^{CP} = -4.8$ kcal/mol.

Note that the aqueous profiles are much flatter than the gas-phase profiles for all species, due to solvent screening. This is illustrated in Figure C.8, which shows the T/BrU FTB profiles in both phases on the same scale.

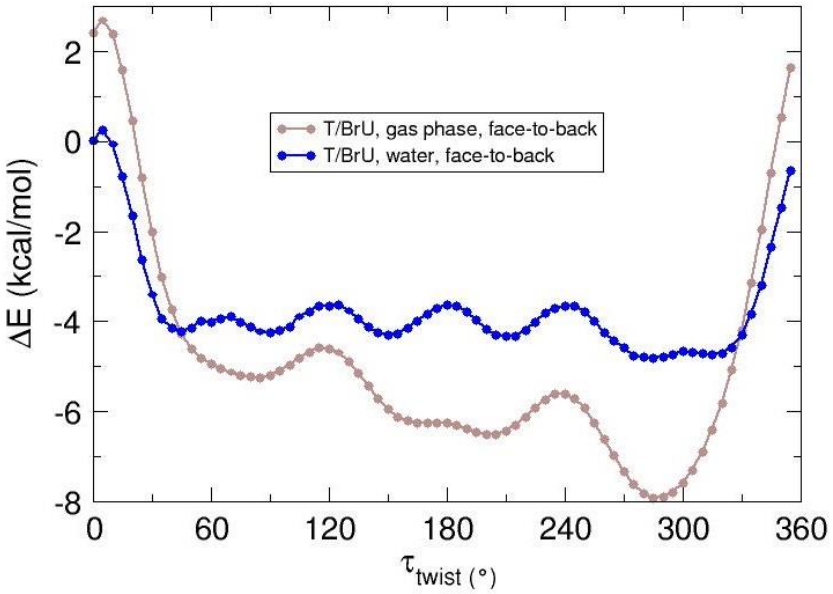


Figure C.8: T/BrU FTB profiles in the gas phase and water shown on the same scale.

Appendix D: Bugg model of BrU-G mispairing

A mechanism for the mispairing of 5-bromouracil in its major tautomeric form was proposed in the 1970s by Bugg *et al.* (see Chapter 5).[73, 77, 78] X-ray diffraction in crystalline 5-halopyrimidines showed that they adopted a rotated conformation (compared to the non-halogenated parent compounds), which allowed the halogen atoms to stack directly above the aromatic rings of adjacent molecules, presumably allowing favourable dispersive and/or electrostatic interactions. Bugg proposed that halogen substituents induced similar rotation in the stacking of 5-halopyrimidines in DNA, and that the resulting orientations allowed favourable H-bonding interactions between BrU and G, but not A.

This hypothesis was based on crystal structures of pure (or co-crystallised) 5-halopyrimidines, not on data from polynucleotides, which was scarce at the time. To investigate whether the proposed geometric changes occur in nucleic acids containing BrU, and whether any such trends could be correlated with stacking energies, we expanded our set of stacked dimers containing BrU from the Protein Data Bank.

Due to the helical twist of DNA, the 5-substituent on a pyrimidine lies closer to the aromatic ring of its 5' neighbour than to that of the 3' neighbour (Figure D.1). The geometric alteration proposed by Bugg, in which 5-halopyrimidines rotate to allow close halogen–ring contacts, is therefore most suited to dimers with the halogenated base on the 3' side. We isolated the geometries of every 5'-X/BrU-3' stacked dimer that we could find in the PDB, yielding 63 such dimers in all. The corresponding gas-phase stacking potential energies were then calculated by the method described in Chapter 5, Section 5.2.3: M06-2X/6-31+G(d) partial optimisations using Gaussian 09, followed by RI-mPW2PLYP-D/aug-cc-pVTZ single-point energies using ORCA 2.8 (dispersion being calculated *via* ORCA's "vdw" keyword).

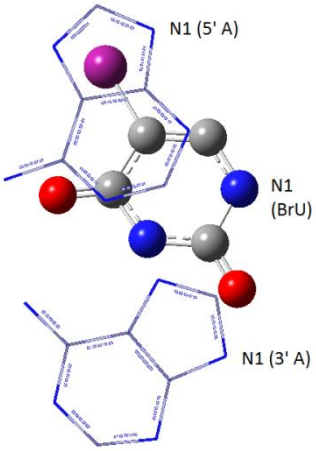


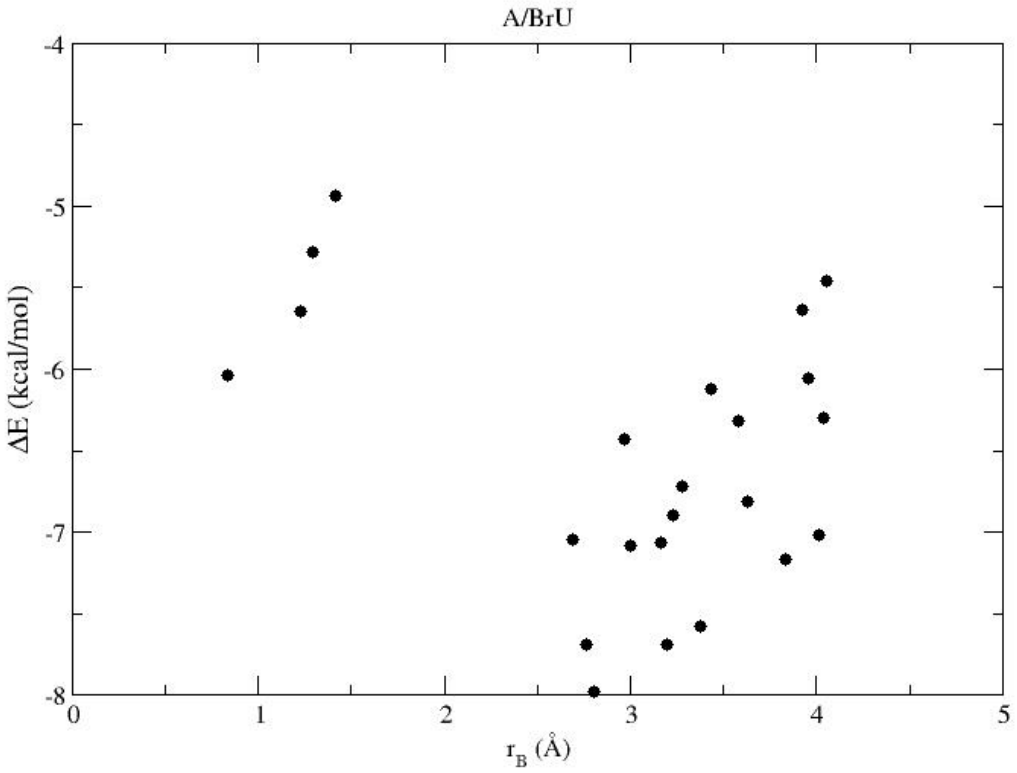
Figure D.1: The Br atom of BrU (purple) stacks closer to its 5' neighbour than its 3' neighbour. Sequence: 5'-A/BrU/A-3' from PDB entry 115D. 5' A shown by wires (top), 3' A shown by wires (bottom), BrU shown as spheres. N1 atoms labelled. H atoms omitted.

We also developed a new module for the *"stack_geometry"* program described in Chapter 5. For any stacked dimer X/BrU, this module takes as its argument the Cartesian coordinates of four atoms: the Br atom of BrU, and the C2, C4 and C6 atoms of X if it is a pyrimidine, or the C2, C4 and C5 atoms of X if it is a purine. The module then calculates the horizontal component, which we denote r_B, of the distance between the Br atom and the midpoint of the six-membered ring of X.

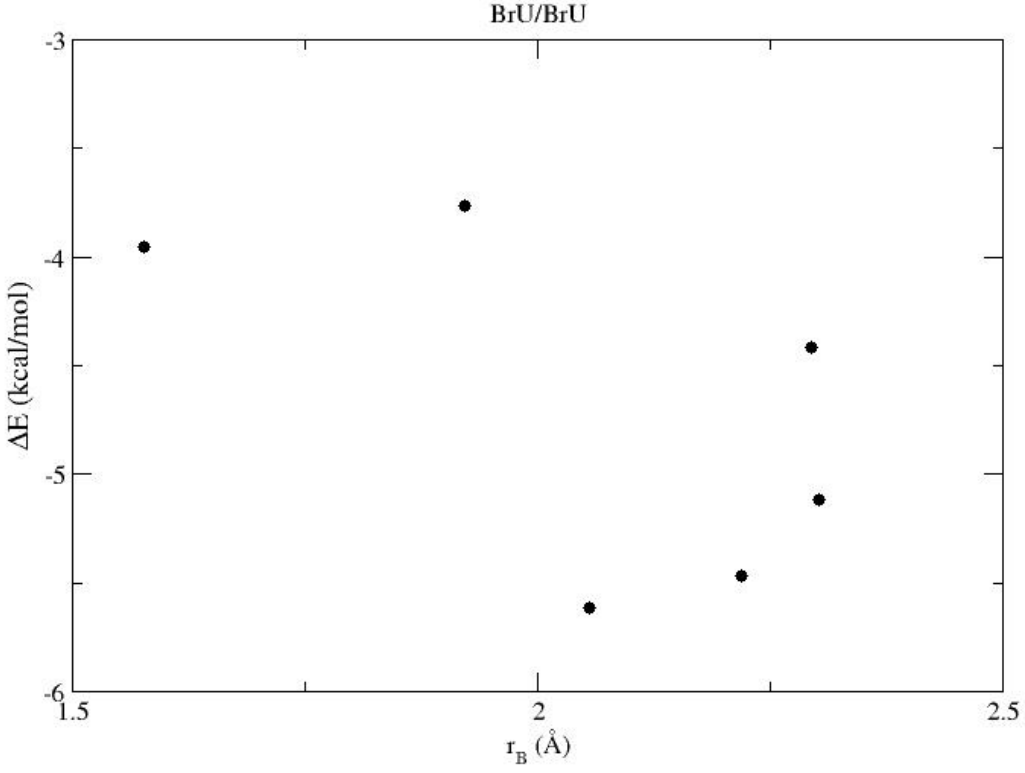
The concept of r_B can be generalised to include X/T dimers, in which the first argument (the location of the 5-substituent on the 3' pyrimidine) refers to the methyl carbon of T rather than the Br atom of BrU. If Bugg's hypothesis is correct, r_B should be shorter in X/BrU than in X/T dimers, reflecting rotational adjustments which allow close contact between the aromatic ring of X and the Br atom of BrU. Moreover, the stacking strength of X/BrU dimers should be negatively correlated with r_B.

Figures D.2–D.6 show, for each type of X/BrU dimer (X = A, BrU, C, G or T), scatter plots of r_B against the corresponding stacking energies.

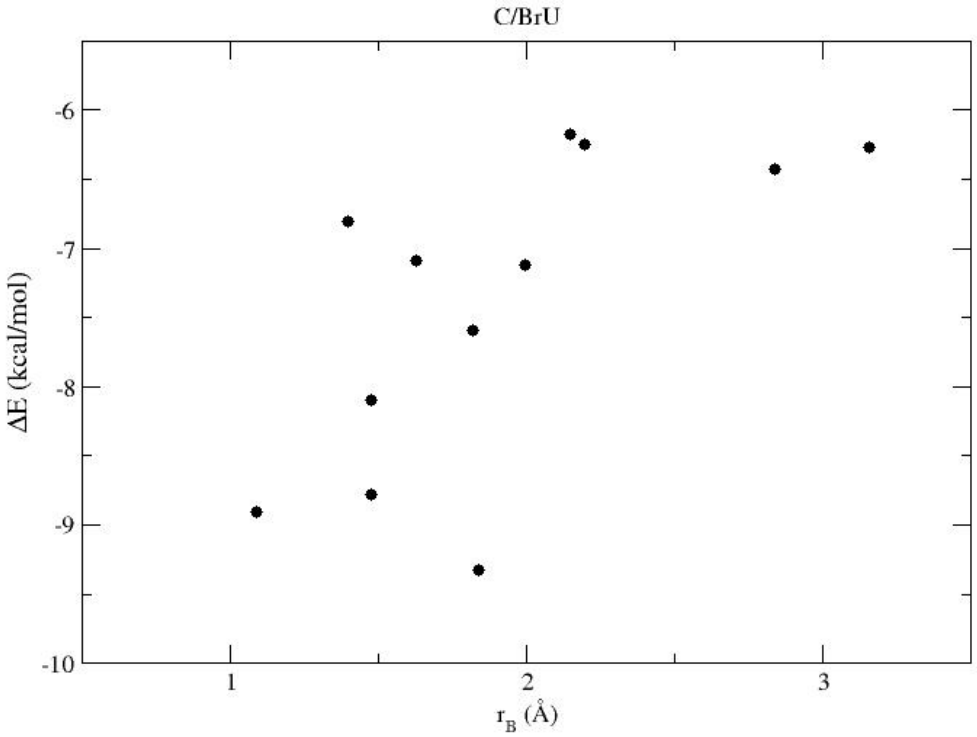
Stacking energy vs. r_B(Br-aromatic horizontal distance)



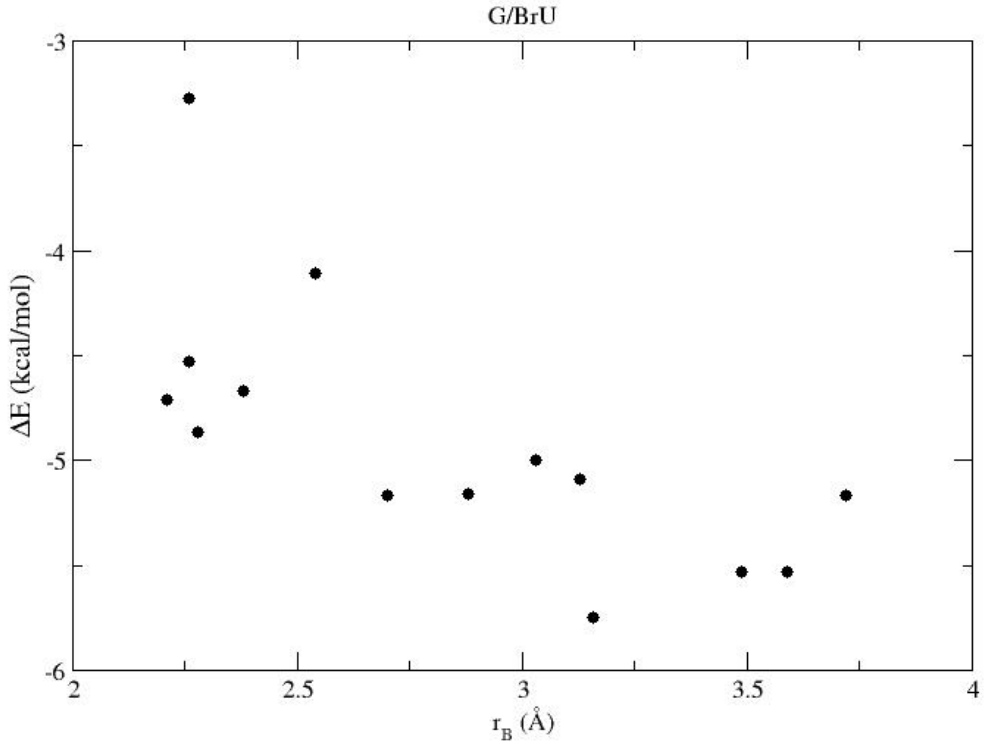
Stacking energy vs. r_B(Br-aromatic horizontal distance)



Stacking energy vs. r_B(Br-aromatic horizontal distance)



Stacking energy vs. r_B(Br-aromatic horizontal distance)



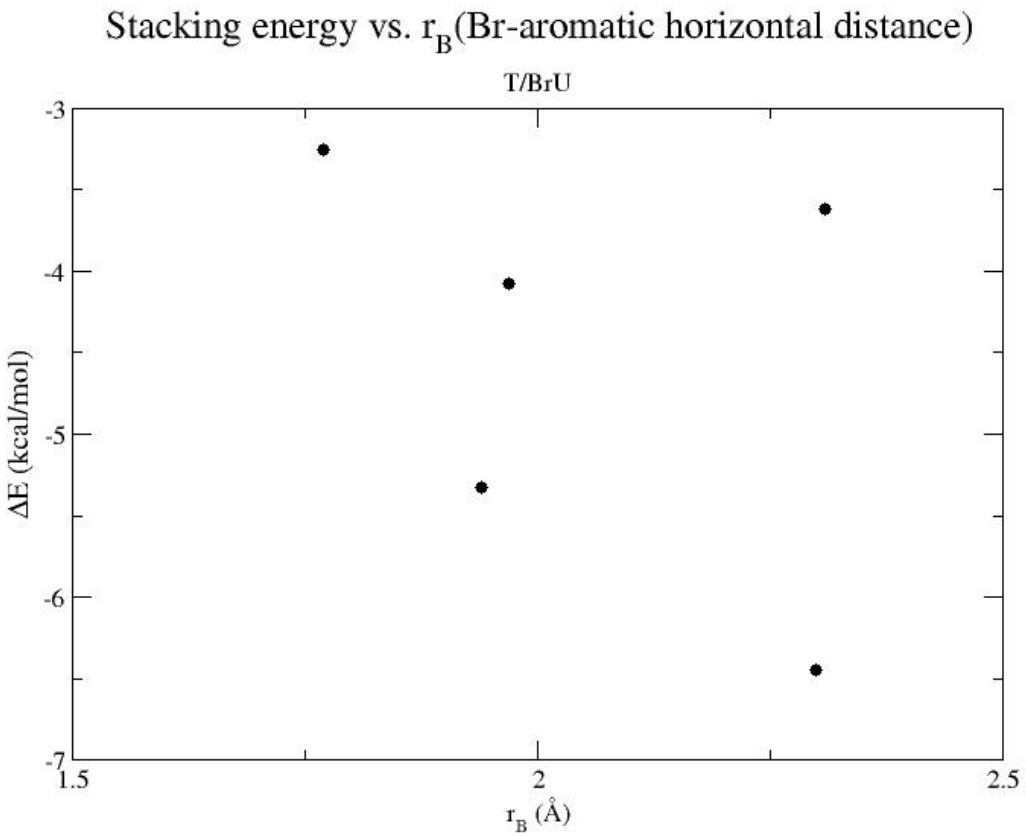


Table D.1 compares the mean values of r_B when the 3' base is BrU or T, separated into five columns according to the 5' base. The BrU-3' data were calculated from the set of 63 X/BrU dimers mentioned above, while the T-3' data were calculated from the 36 X/T dimers among the T-containing dimers used in Chapter 5 (listed in Appendix B). The mean r_B is shorter in A/BrU than in A/T, as per Bugg's prediction, but when X = BrU, C or T, the mean r_B is actually longer for BrU-3' than for T-3', and when X = G, the mean r_B is identical for BrU-3' and T-3'. In other words, the bromine atom on intra-helical BrU does *not*, on average, form closer contacts with the aromatic ring of the 5' neighbour than the methyl carbon on intra-helical T does. The data in Table D1 therefore offer no support to Bugg's model.

	r_B, 5'-A	r_B, 5'-BrU	r_B, 5'-C	r_B, 5'-G	r_B, 5'-T
BrU-3'	3.0	2.1	1.6	2.8	2.1
T-3'	3.3	1.8	1.3	2.8	1.4

Table D.1. Mean r_B (horizontal component of the distance from either the Br atom of BrU or the methyl carbon atom of T to the midpoint of the six-membered ring of the 5' neighbouring base, in Å) when the 5' neighbour is A, BrU, C, G or T.

Appendix E: Base flipping of 2-aminopurine dinucleosides

E.1. Background

2-Aminopurine (2AP) is a structural analogue of adenine, with the amino group at the 2-position rather than the 6-position. It can be incorporated into DNA in place of adenine, and base-pairs with thymine[vi] in Watson–Crick geometry (although the second hydrogen bond is between N2H(2AP) and O2(T) rather than N6H(A) and O4(T)). The incorporation of 2AP is believed to cause only slight changes in overall DNA structure.[221-223] But unlike the natural bases – which relax from excited states by radiationless processes – 2AP is highly fluorescent: in solution, it has a high quantum yield (~0.7) and long fluorescence lifetime (~10–12 ns). Furthermore, its absorption wavelength (300–320 nm) is considerably red-shifted from that of the natural bases (~260 nm), so it does not couple with the neighbouring bases in DNA, and can be selectively excited even in long duplexes.[222, 224, 225] However, its quantum yield in the duplex, as well as in dinucleotides, is heavily quenched compared to solution,[222, 224-226] and this quenching is sequence-dependent.[226] Furthermore, 2AP in the duplex exhibits a long-wavelength emission band that is absent in solution, as well as the more familiar short-wavelength band.[226] Due to this combination of environment-sensitive fluorescence and minimal structural perturbation, 2AP has been widely used as a fluorescent probe of local dynamics in DNA.[221-223, 226, 227]

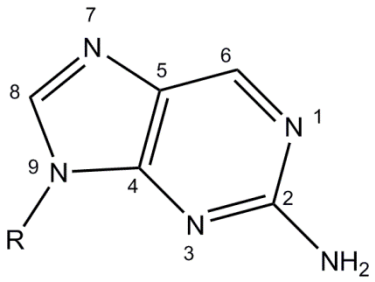


Figure E.1: 2-Aminopurine.

One phenomenon that has been explored in this way is *base flipping*. Various enzymes exist whose function is to unstack a specific target base in DNA, by rotating it fully out of the duplex and into a catalytic pocket of the enzyme. This behaviour was first observed in various methyltransferase enzymes, which selectively flip and then methylate bases as part of certain processes within the cell. Base flipping without subsequent methylation also plays a role in DNA recognition.[223, 227] Base flipping obviously causes a significant perturbation to the local environment experienced by bases – especially for the target base itself, but also for the stacking of proximal bases, due to the creation of an abasic (unoccupied) site and subsequent relaxation around that site. Evidence for base flipping can therefore be inferred from changes in the fluorescence of 2AP (generally an increase in emission intensity) when it is present either as the target base or as a proximal base in a stretch of DNA exposed to a base flipping enzyme.

While 2AP in solution fluoresces with a single, characteristically long decay lifetime, its emission behaviour in the duplex is more complicated, with typically four lifetimes spanning two orders of

[vi] 2AP can also mispair with cytosine, causing transition mutations in the same way as the A*-C mispair, but this will not be discussed here.

magnitude.[222, 225, 226, 228] While the shortest lifetime corresponds to the fully stacked, quenched state, and the longest presumably represents an unstacked state similar to the one obtained by base flipping, it has been harder to account for the origin of the other two.[222] One possibility is that they arise from partially stacked intermediate states, but these are difficult to verify experimentally, and it is here that calculations may play a role. One difficulty is that the quenching mechanisms are not fully understood.[225] The existence of multiple fluorescence lifetimes implies that the excited dinucleotide occupies several states, but it is not clear whether these states arise before or after excitation. The former case relies upon a "static" interpretation of 2AP fluorescence, in which each lifetime corresponds to a ground-state conformation of the dinucleotide, while the latter implies a "dynamic" interpretation, in which one or both of the intermediate-lifetime states is formed after excitation, either directly or by radiationless relaxation from a higher excited state. Interestingly, one early study using time-dependent density functional theory (TD-DFT) predicted that both could occur, depending on sequence: the quenching in 2AP/Pu dinucleotides (Pu = purine) should be static, owing to significant electron delocalisation in the absorbing HOMO of 2AP as a result of stacking; but the quenching in 2AP/Py dinucleotides (Py = pyrimidine) should be dynamic, and was attributed to internal conversion from the S_2 excited state to an S_1 dark state that is much lower-lying in the dinucleotide than in free 2AP.[228]

Conceivably, the excited states responsible for the intermediate fluorescence lifetimes of aqueous 2AP/N might arise through photoexcitation of ground-state conformations that are structurally intermediate between the fully stacked and fully open minima. However, little has been predicted in detail about the structures or thermodynamics of these putative states. In particular, it is not clear whether they should correspond to discrete minima on the 2AP/N potential energy surface, with specific geometries and equilibrium populations, or if they more likely represent the average of a continuum of states along a smooth region of the hypersurface between the stacked and open minima. If they are discrete minima, what are their relative energies, and how high are the barriers to interconversion?

E.2. Methodology

The work described in this Appendix was carried out in collaboration with Dr. Darren Smith and Prof. Anita Jones of the University of Edinburgh. Dr. Smith obtained the PDB structures, designed the LIIC procedure for 2AP/T, and carried out the single-point calculations on 2AP/T. Prof. Jones has been involved throughout in planning the overall strategy and interpreting the results.

Our aim was to search for energy-minimum structures of aqueous 2AP/N and A/N dinucleosides in geometries intermediate between a conventional stacked structure and an "open" structure, mimicking the pathway of enzymatic base flipping. A complete search of the conformational space of these systems would be impossible; instead, we decided to search along a hypothetical linear pathway for base flipping, using the Linear Interpolation of Internal Coordinates method (LIIC)[229] to generate candidate intermediate structures.

The atomic coordinates of a stacked 5'-A/T-3' dinucleoside within a DNA helix were taken from PDB entry 3R86 (the crystal structure of a regular B-DNA decamer containing only natural bases). We isolated a molecular fragment from the helix, consisting of two bases, two deoxyribose units (terminating in C5'-$O_{hydroxyl}$ at the 5' end, and $O_{hydroxyl}$ at the 3' end), and a bridging phosphate group. This defines it as a deoxydinucleoside monophosphate, a unit of DNA structure which has been investigated in previous computational studies.[230-233] Since H atoms are not present in PDB structures,

we saturated the bases and deoxyribose units by adding H atoms, at positions corresponding to the canonical tautomers in DNA. We also neutralised the overall molecular charge, by adding an H atom to one of the non-bridging O atoms of the phosphate group. We then mutated the A monomer to 2AP, by swapping the amino group at C6 with the H atom at C2. Finally, the resulting stacked 2AP/T dinucleoside was optimised at the M06-2X/6-31+G(d) level of theory, in aqueous solution modelled by the SCRF-IEFPCM method, using Gaussian 09.[136] (All subsequent calculations in this chapter employed the same model chemistry, except where stated otherwise.)

The atomic coordinates of a flipped 5'-2AP/T-3' dinucleoside were taken from PDB entry 2IBS. This entry is the crystal structure of a methyltransferase enzyme, M. TaqI, which targets adenine for base flipping.[223] The crystalline enzyme is complexed with a DNA decamer containing 2AP in place of A at the target site for base flipping, and this 2AP residue is flipped out of the helix in a similar conformation to that usually observed in the natural target. We saturated the flipped 2AP/T dinucleoside with H atoms, then geometry-optimised it, in the same way as for the stacked structure.

To generate intermediate structures for 2AP/T, the two optimised structures described above served as the start and end points, respectively. For the other three dinucleoside systems that we considered (2AP/G, A/T and A/G), the start and end points for the generation of intermediate structures were obtained from those of 2AP/T as follows.

For 2AP/G, the stacked structure was created by mutating T to G in stacked 2AP/T, then optimising it. The flipped structure was created by mutating T to A in flipped 2AP/T, optimising it, then mutating A to G and optimising it again. For A/T, the stacked and flipped structures were created by mutating 2AP to A in the optimised stacked and flipped structures, respectively, of 2AP/T, and optimising them. For A/G, the stacked and flipped structures were created by mutating 2AP to A in the stacked and flipped structures, respectively, of 2AP/G, and optimising them.

The optimised stacked 2AP/T and 2AP/G structures contain a hydrogen bond between NH_{amino} (2AP) and the deoxyribose O atom on the 3' side. This hydrogen bond is absent from the A/T and A/G minima, due to the different location of the amino group in adenine.

For each of the four dinucleoside systems, the LIIC was then implemented as follows. Firstly, we used Molden to write Z-matrices of the start and end points. We ensured that both Z-matrices used the same set of internal coordinates (i.e. the same combinations of atoms were used to define each distance, angle and dihedral). Then, for each system, we interpolated a series of 19 structures intermediate between the start and end points, by sequentially adjusting the value of each coordinate in the range between its start and end values. This adjustment was done in increments of $x/20$, where x is the difference between the start and end values of each coordinate. This resulted in 21 structures for each dinucleoside system (including the start and end points), representing one possible continuous transition pathway from the stacked structures to the corresponding flipped structures. Figures E.2 and E.3 show the pathways for 2AP/T and 2AP/G.

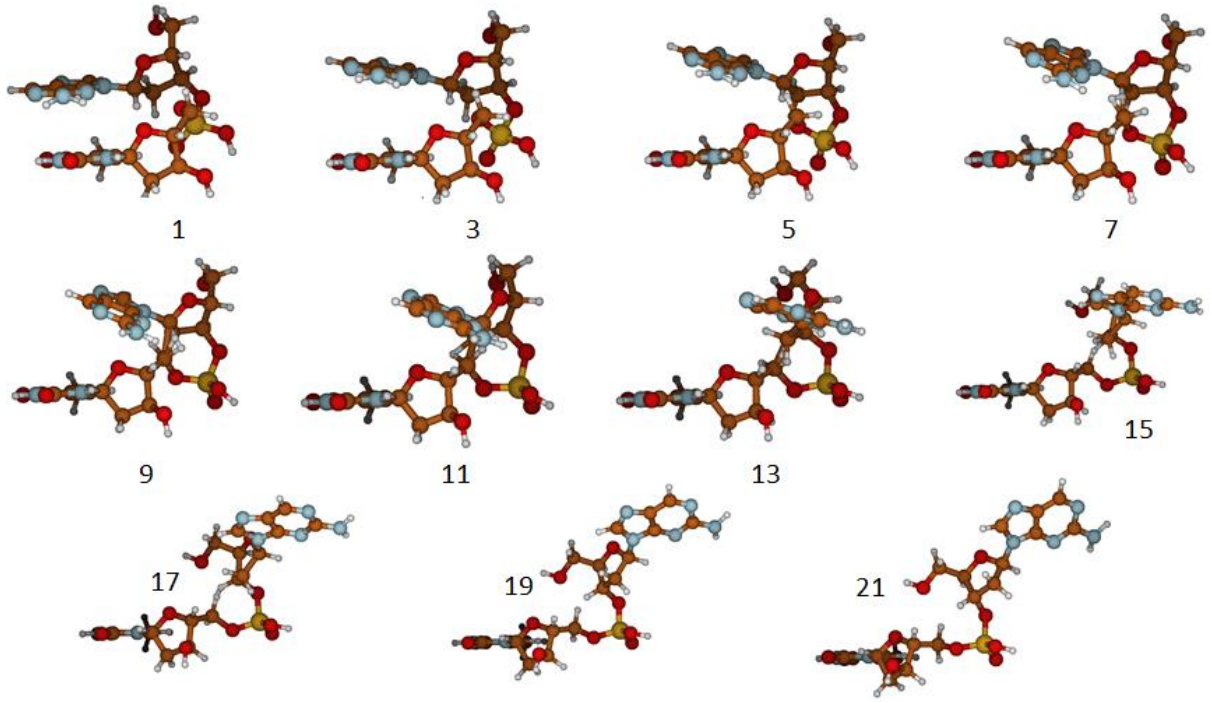


Figure E.2: The LIIC "flipping" pathway for 2AP/T (odd-numbered points only).

Two important caveats must be noted, both of which relate to the fact that this is an arbitrary way of creating intermediate structures, since there is no "unique" or optimal set of intermediate points.

Firstly, the choice of internal coordinates, which determines the pathway of the intermediates, is to some extent arbitrary.

Secondly, each angle and dihedral can be rotated in two directions ("clockwise" or "anti-clockwise"). The difference, x, between the start and end values has two possible magnitudes, one less than and one greater than 180°. For example, a rotation from 100° to 150° could be accomplished by 20 increments of +2.5° (x = 50°), but it could also be accomplished by 20 increments of −15.5° (x = −310°), since 100° − 310° = −210° ≡ 150°. For coordinates that change only slightly during the transition (e.g. those maintaining the planarity of the nucleobases), one direction requires much smaller steps than the other, and is obviously preferable. However, for coordinates that undergo larger changes (e.g. backbone dihedrals), both directions require significant step sizes, and a choice must be made. For each angle and dihedral we have, by default, chosen the direction with the smaller value of x, hence the smaller step size. However, one backbone dihedral in each system had to be rotated through the larger value of x, because rotation through the smaller value yielded intermediate geometries containing atomic clashes. This dihedral was, in each system, defined by the atoms O_{bridge}(5')-P-O_{bridge}(3')-C5'(3'), where O_{bridge} are the bridging atoms in the phosphate group, and C5'(3') is the deoxyribose C5' atom on the 3' side (towards either T or G).

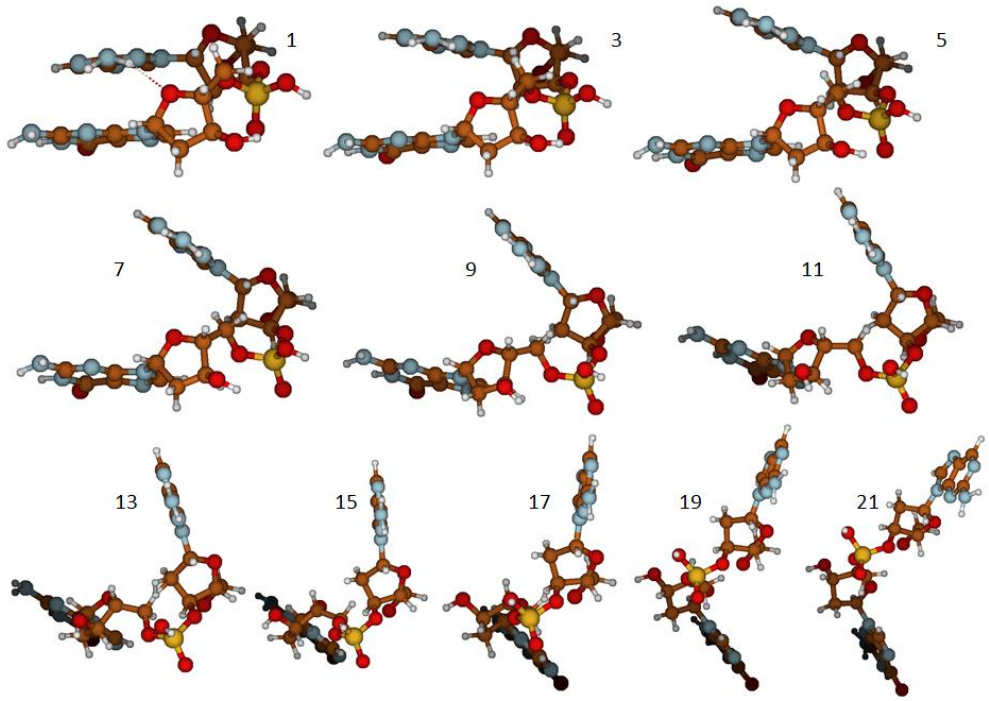


Figure E.3: The LIIC "flipping" pathway for 2AP/G (odd-numbered points only).

When viewed as animated sequences, the pathways begin with an essentially vertical separation of the stacked bases, accompanied by an increase in the angle between their planes, followed by a rotation of the backbone such that the bases move apart horizontally. Despite the change of dihedral rotational direction mentioned above, some unrealistically close contacts still occur between hydrogens on the two deoxyribose groups for Points 6 to 18.

Having obtained the intermediate structures in this way, the procedure was as follows: launch geometry optimisations using each intermediate as the starting structure, then search for transition states between the resulting minima.

All calculations were carried out using Gaussian 09.

To discourage the formation of hydrogen bonds whose equivalents would not be possible in a full DNA helix, the H atom on the phosphate group was removed from each structure before the optimisation. The calculations were then carried out with a charge of −1 and a multiplicity of +1, i.e. the phosphate groups were deprotonated, but the electron associated with the H atom remained, giving an anionic, closed-shell system. Churchill and Wetmore found that this protocol reproduced stacked dinucleoside geometries much more satisfactorily than neutral, protonated phosphates when using M06-2X.[232]

In practice, we did not have to use all the intermediates. Anticipating that several sequential points might sometimes optimise to the same minimum, we began by staggering the choice of starting structures (e.g. following Point 1 with Point 5), then returning to fill in the gaps where necessary. Whenever non-sequential points were found to yield the same minimum, we assumed that the points in between would do so too, and by neglecting them we saved considerable computer time.

E.3. Results

The results for each of the four systems will be presented in turn.

E.3.1. 2AP/T

For 2AP/T, our first action was to calculate single-point energies of all 21 LIIC-generated structures. However, in two of these structures (Points 11 and 12), one pair of atoms (H27 and C36 in the Z-matrix) were too close together for an SCF calculation to be possible. The SCF potential energies of the other 19 structures are shown in Figure E.4. The fully stacked structure (Point 1) is the lowest in energy, while the fully extended structure (Point 21) lies 23 kJ/mol above it. With the exception of Point 2, all the intermediate structures lie energetically above both the fully stacked and the fully extended structure. The least stable structure for which SCF convergence was possible, Point 13, lies nearly 3500 kJ/mol above Point 1, indicating that this geometry is not physically plausible. In fact, all the structures from Points 6 to 18 lie more than 200 kJ/mol above Point 1. However, the results do confirm that converged SCF energies can be obtained for most of the intermediate structures generated by our methodology.

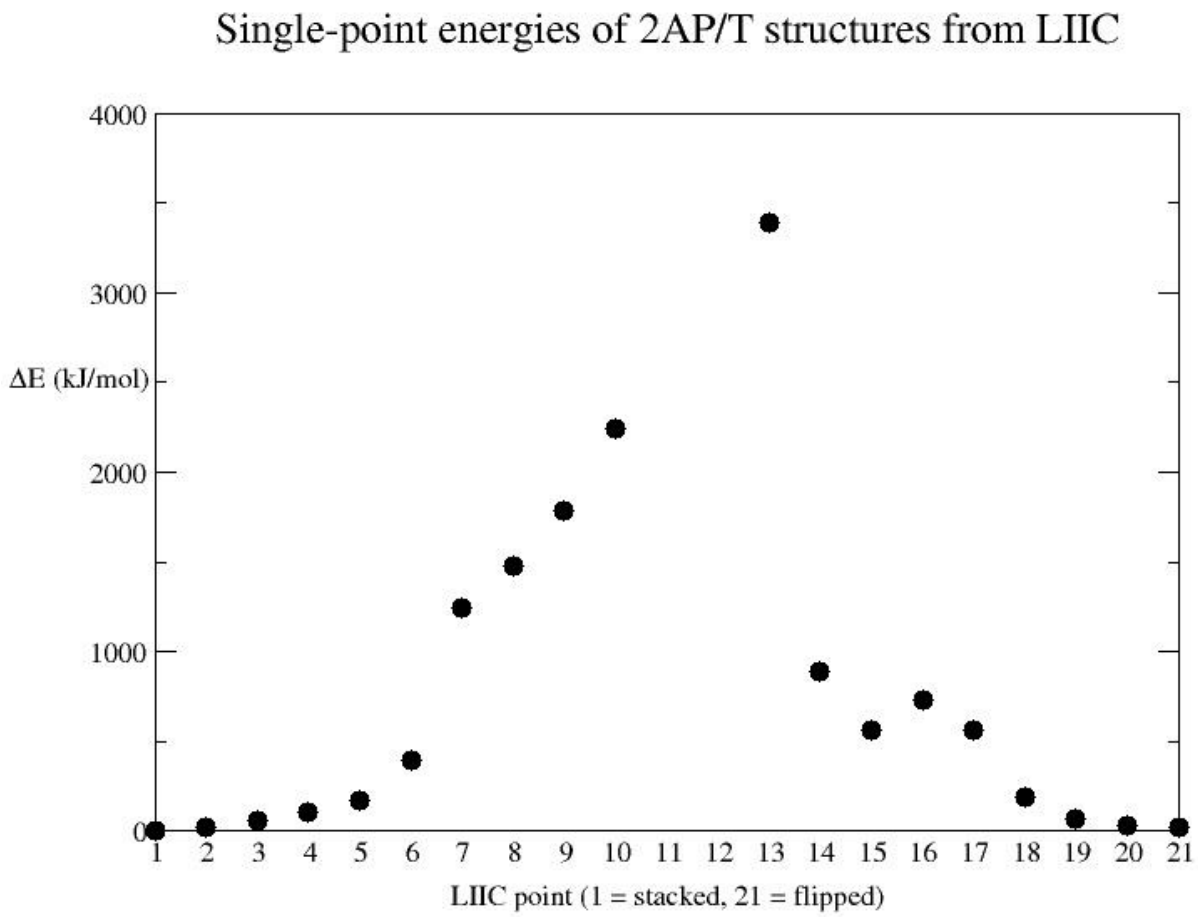


Figure E.4: M06-2X/6-31+G(d) single-point energies of 19 of the 21 LIIC points of 2AP/T, relative to Point 1.

Nine distinct 2AP/T minima were obtained from the initial set of geometry optimisations. Several optimisations resulted in significant structural changes, and two required the use of either "Opt=CalcFC" or "Opt=XQC" to successfully locate minima. Points 11 and 12 could not be

geometry-optimised, owing to the atomic clash mentioned above. In contrast to the single-point energies shown in Figure E.4, four of these nine minima lie *below* Point 1 in energy, with Point 6 being the lowest. Unlike the single-points, none of the intermediate points are hugely (i.e. more than 100 kJ/mol) less stable than the endpoints, indicating that the very high relative energies in Figure E.4 arose from atomic clashes in the unrelaxed LIIC geometries, rather than inherent instability of intermediate geometries.

We then searched for transition states linking "consecutive" minima, i.e. those minima which arose from optimising consecutive points (or consecutive groups of points that converged to the same geometry). In some cases we also considered non-consecutive pairs of minima. The transition state searches were initially carried out using the QST2 method, whereby the optimised structures of the two minima served as the two input geometries. In some cases this method successfully located a transition state, as confirmed by the presence of one imaginary harmonic vibrational frequency. Where QST2 failed to locate a turning point, the search was resumed by one of two methods: either a QST3 search was launched, using the final optimisation step from QST2 as the guess TS geometry along with the original input minima, or a conventional "Opt=TS" search was launched, using only the final QST2 geometry as the guess. In the latter case, the "Opt=NoEigenTest" option was also invoked, which suppresses testing the curvature of the potential energy surface, and a separate frequency calculation was performed on the final optimised structure. This combination of methods successfully yielded transition states for all pairs of input minima, except the pair of Points 9 and 10.

In five cases, however, the resulting transition state was *lower* in SCF energy than one, or both, of the minima that were originally used to locate it, implying that it did not actually connect those two minima on the potential energy surface of 2AP/T. We therefore "followed" the imaginary frequencies (cf. Chapter 5, Section 5.3.1) of all transition states located, to determine whether any of them actually connected the minima they were expected to. Each of these pairs of optimisations (following the vibrations in opposite directions) yielded two different minima, except for the optimisations based on the TS apparently linking Point 1 and Point 6, which both yielded the same minimum. However, several of the resulting minima were different from any that had been obtained from the LIIC intermediates, and in not a single case did a given pair of optimisations yield both the expected minima.

Therefore, after this second round of geometry optimisations, a total of 25 distinct 2AP/T energy minima had been located. This includes two which were optimised from starting structures in which Points 1 and 9, respectively, were modified by rotating the 3' hydroxyl group. The 25 structures are shown in Figure E.5, in order of increasing SCF energy (ΔE) relative to the lowest minimum. That minimum, here labelled "7 to 8 −", is partially unstacked (heavily tilted) and contains two hydrogen bonds. It was obtained by geometry-optimising the structure created by subtracting (hence the minus sign), from each atomic coordinate, the displacements of the imaginary vibrational frequency of the transition state that was expected to link the structures obtained from Points 7 and 8. That transition state, in turn, had been obtained by a QST3 optimisation which used, as its guess geometry, the final structure from the attempted QST2 search that used the optimised Points 7 and 8 as inputs. Similar sequences of calculations lie behind all the other minima labelled with "+" or "−".

Figure E.5 also displays the estimated Gibbs free energies (ΔG) of the minima at 298.15 K. These energies are given relative to "7 to 8 −", but note that the structure with the lowest Gibbs free energy is "8 to 9 −", which is geometrically quite different from "7 to 8 −", being intermediate between

stacked and flipped. While the potential energies span a range of 57 kJ/mol, the free energies span only 33 kJ/mol.

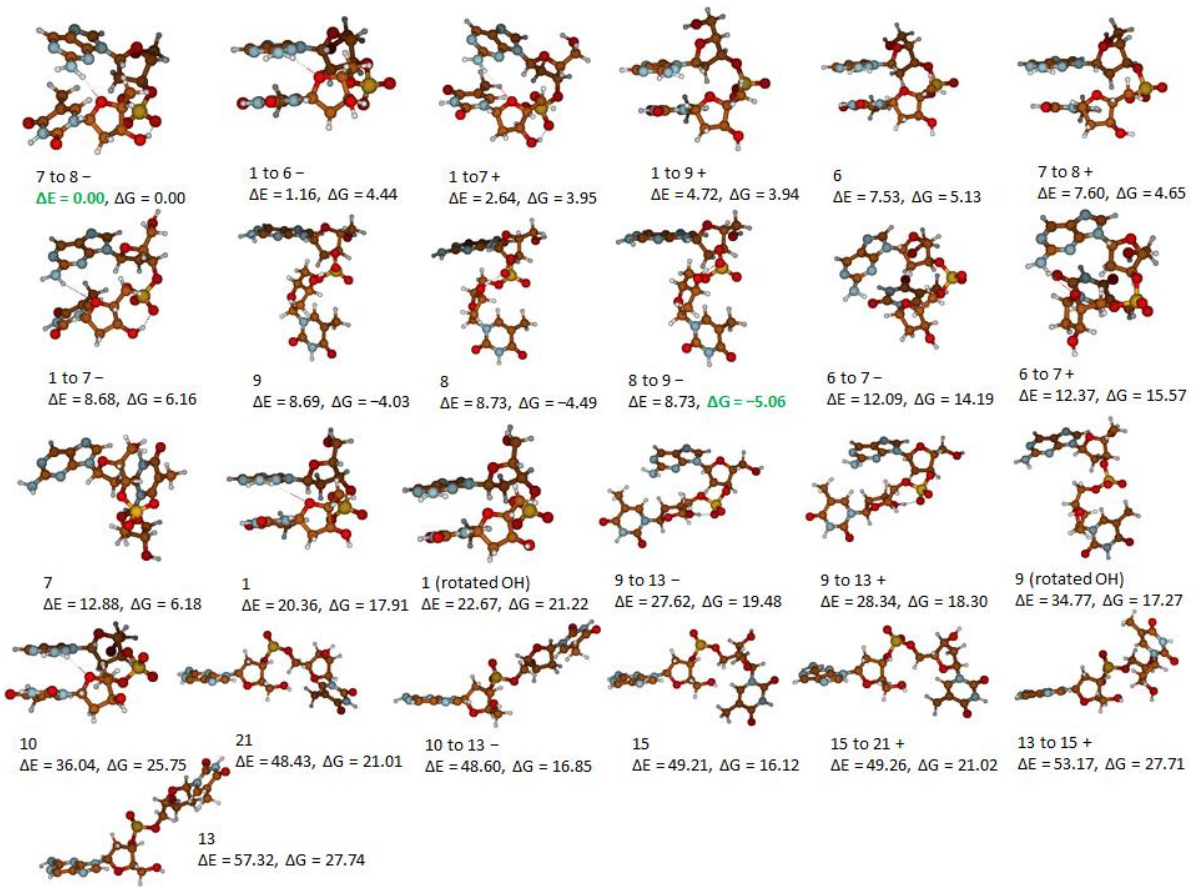


Figure E.5: 25 minima of 2AP/T, ordered by potential energy. Also shown are the free energies, relative to the free energy of the potential energy minimum. Units: kJ/mol.

By obtaining the minima that are actually connected by the transition states, we were able to calculate the energetic barriers to conversion between these pairs of minima (though not between any other pairs). These barriers are given in Table E.1. In SCF energies, they range from 0.2 to 25.4 kJ/mol, while the free energy barriers range from 1.5 to 24.6 kJ/mol. One of the largest barriers separates the tilted structure "7 to 8 −" from the stacked structure "7 to 8 +". Otherwise, however, these barriers do not relate to conversion between geometrically very distinct structures.

Pair of minima	Barrier (E_a), left-to-right	Barrier (E_a), right-to-left	Barrier (G_a), left-to-right	Barrier (G_a), right-to-left
1-to-7 + ←→ 1-to-7 −	8.29	2.24	4.70	2.48
1-to-9 + ←→ 6	7.18	4.38	5.22	4.03
6-to-7 + ←→ 6-to-7 −	0.41	0.69	1.46	2.84
7-to-8 + ←→ 7-to-8 −	10.51	18.11	18.44	23.09
9 ←→ 8-to-9 −	6.94	6.90	6.35	7.38
9-to-13 + ←→ 9-to-13 −	2.33	3.06	5.27	4.09
13 ←→ 10-to-13 −	16.71	25.43	13.77	24.64
13-to-15 + ←→ 21	9.81	14.55	9.67	16.37
15-to-21 + ←→ 15	0.22	0.27	5.25	10.15

Table E.1. Potential energy (E_a) and free energy (G_a) barriers between some 2AP/T minima (kJ/mol).

E.3.2. 2AP/G

Six distinct minima were obtained from the initial round of optimisations. The lowest SCF energy belongs to the structure obtained by optimising Points 1–4, i.e. the stacked starting point of the LIIC. This structure is labelled "Point 1". Point 12 could not be optimised due to a close interatomic distance.

We searched for transition states between consecutive minima, in the same way as for 2AP/T, and successfully located five. However, two of these had lower SCF energies than one of the minima they were expected to connect. By "following" the imaginary frequencies of all the transition states, we located a further six distinct minima. For only one transition state did this procedure return both of the original minima that it was expected to connect (Points 13 and 14).

Therefore, in total, 12 distinct 2AP/G minima were located, and they are shown in Figure E.6. The lowest SCF energy belongs to Point 1, but the lowest Gibbs free energy belongs to the structure obtained from Points 5–10 (labelled "Point 5"), an intermediate structure between stacked and flipped. The potential energies span 54 kJ/mol while the free energies span 32 kJ/mol.

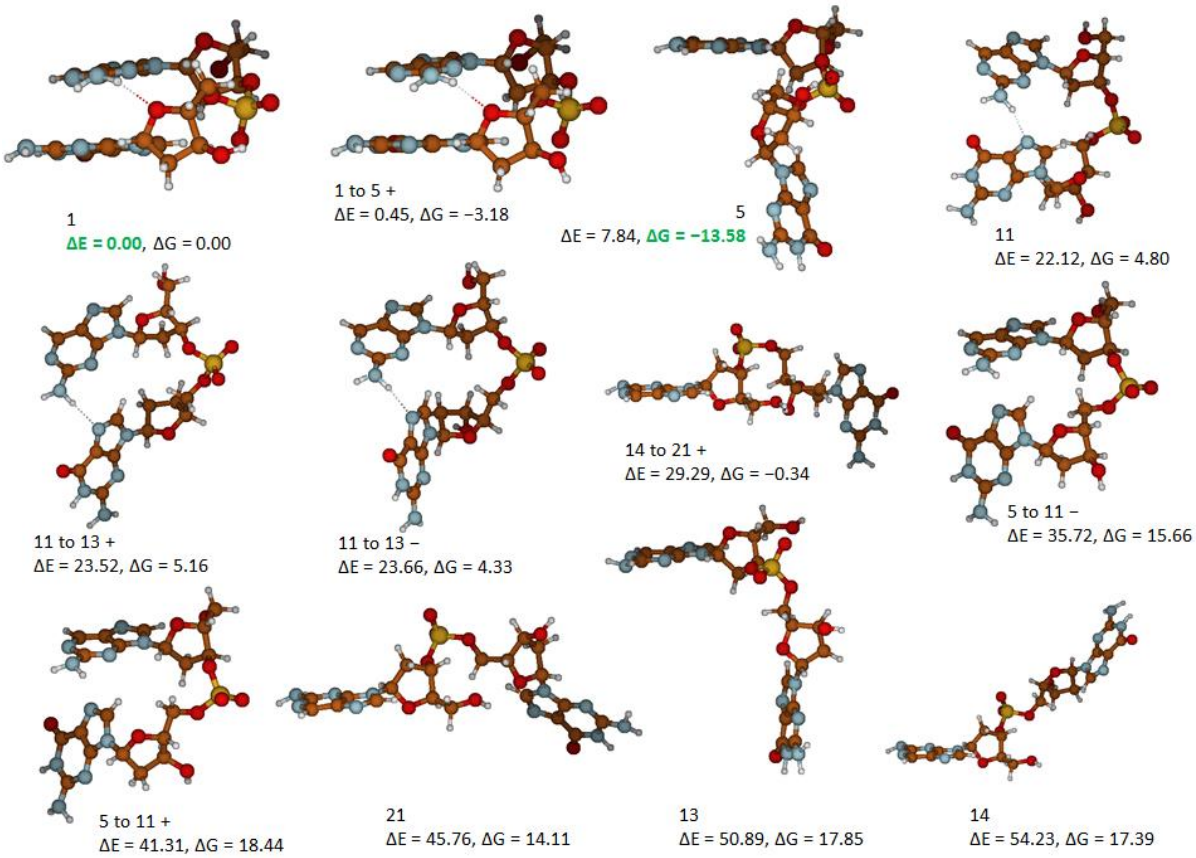


Figure E.6: 12 minima of 2AP/G, ordered by potential energy. Also shown are the free energies, relative to the free energy of the potential energy minimum. Units: kJ/mol.

The barriers (Table E.2) range from 0.1 to 30.2 kJ/mol in potential energy, and from −2.3 to 26.1 kJ/mol in free energy. The one negative free energy barrier (for conversion from Point 1 to "1 to 5 +") may be due to inaccuracies in the calculation of S_{vib}, or it may indicate that Point 1 is not a minimum on the free energy surface of 2AP/G – note that "1 to 5 +" lies 3.2 kJ/mol *below* Point 1 in free energy. In any case, those two structures are similar. The two sets of barriers which pertain to significant structural changes are the ones between Points 13 and 14 (17.4 and 14.1 kJ/mol in potential energy), and the ones between Point 14 and "14 to 21 +" (30.2 and 5.3 kJ/mol in potential energy).

Pair of minima	Barrier (E_a), left-to-right	Barrier (E_a), right-to-left	Barrier (G_a), left-to-right	Barrier (G_a), right-to-left
1-to-5 + ←→ 1	1.19	1.65	0.89	−2.29
5-to-11 + ←→ 5-to-11 −	3.98	9.57	11.59	14.37
11-to-13 + ←→ 11-to-13 −	0.19	0.06	2.06	2.89
13 ←→ 14	17.38	14.05	17.21	17.79
14-to-21 + ←→ 14	30.20	5.27	26.12	8.39

Table E.2. Potential energy (E_a) and free energy (G_a) barriers between some 2AP/G minima (kJ/mol).

At this stage, fewer minima had been found for 2AP/G (12) than for 2AP/T (25). This could have been due to an intrinsic difference between Pu/Pu and Pu/Py dinucleosides, or it could merely have been an artefact of the two different LIIC pathways. To address this, our group searched for

additional 2AP/G minima, by optimising structures that we obtained by mutating T to G in the 2AP/T minima. These calculations were performed by Rebecca Muller Alcalay. This procedure uncovered 15 additional 2AP/G minima, giving 27 in total, all shown in Figure E.7. The lowest SCF energy structure is now a distorted stacked structure obtained from 2AP/T "1 to 6−", while the lowest free energy structure is still Point 5. It therefore appears that the structural diversity of 2AP/T also applies to 2AP/G.

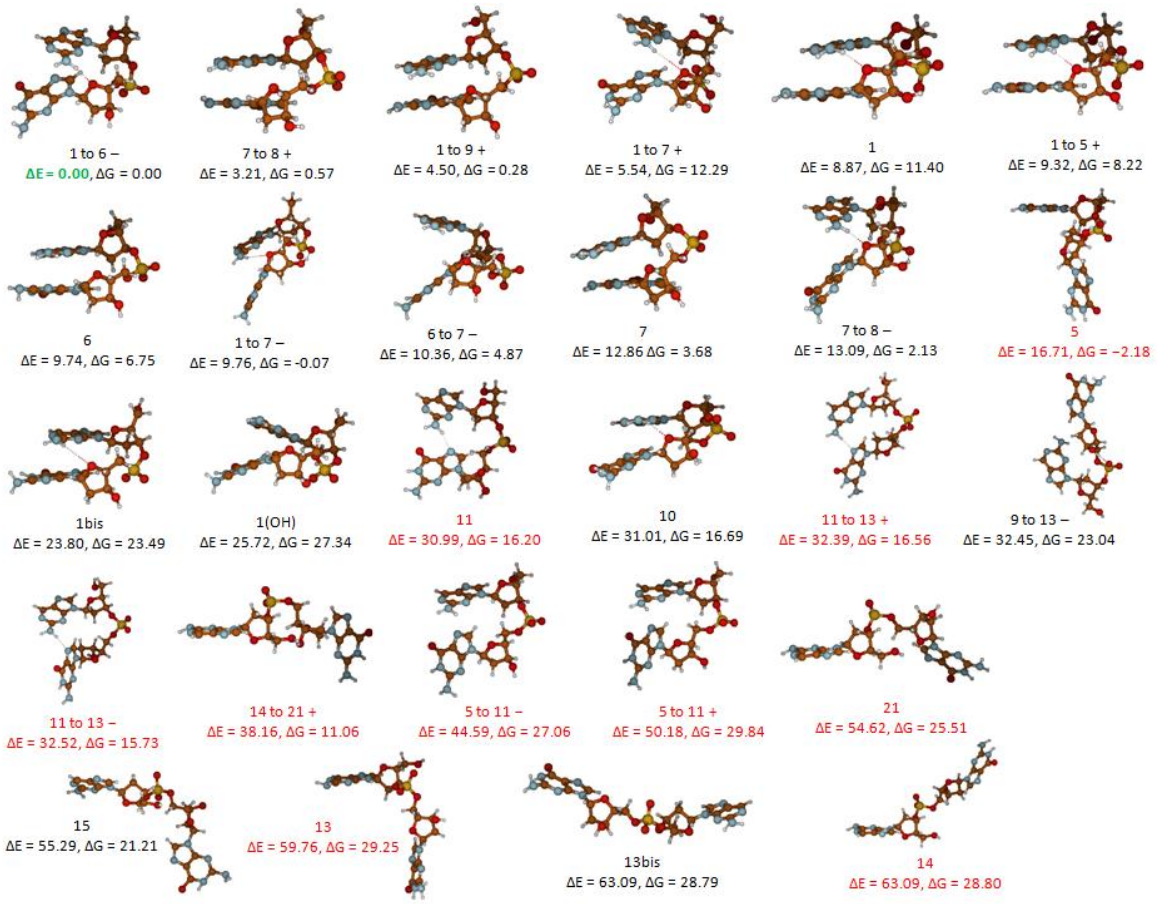


Figure E.7: 27 minima of 2AP/G (including those obtained by mutating 2AP/T minima), ordered by potential energy. Also shown are the free energies, relative to the free energy of the potential energy minimum. Units: kJ/mol.

E.3.3. A/T

Nine distinct minima were obtained from the initial round of optimisations of A/T. The lowest SCF energy belongs to Point 8, an intermediate structure between stacked and flipped. Points 11 and 12 could not be optimised due to a close interatomic distance.

We searched for transition states between consecutive minima in the same way as above. We did not search for a TS between Points 8 and 9, as we deemed them too similar, but we did search for a TS between Points 1 and 8. For the TS between Points 8 and 10, we carried out both QST3 and Opt=TS searches following the unsuccessful QST2 search, and these converged to two different transition states. In total, we obtained nine transition states, one of which had a lower SCF energy than both the minima it was expected to connect. By "following" the imaginary frequencies of all the transition states, we located a further 13 distinct minima. In not a single case did a given pair of these optimisations return both of the expected original minima.

Therefore, in total, 22 A/T minima were located, and they are shown in Figure E.8. The lowest SCF energy belongs to "1 to 6 −", a stacked structure, but the lowest Gibbs free energy is that of Point 9, an intermediate structure between stacked and flipped. The potential energies span 51 kJ/mol while the free energies span 33 kJ/mol.

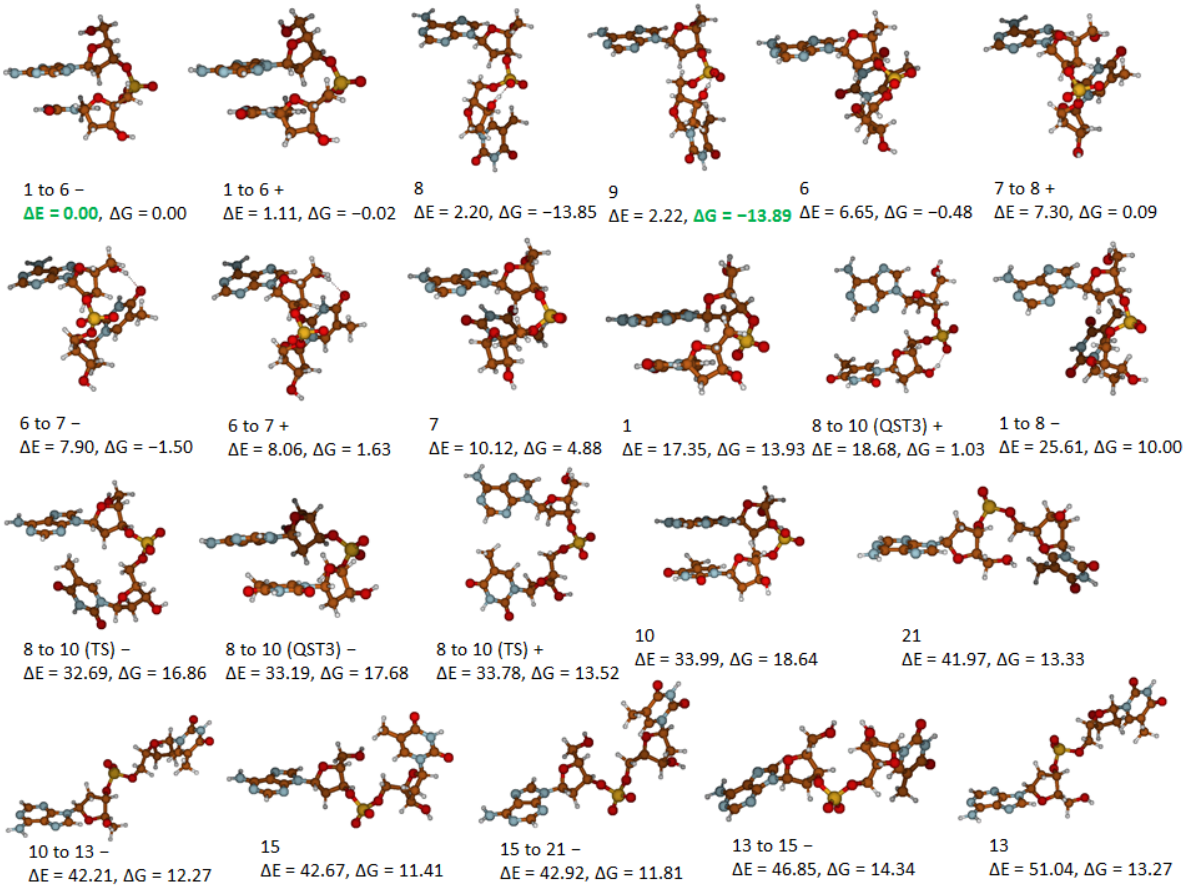


Figure E.8: 22 minima of A/T, ordered by potential energy. Also shown are the free energies, relative to the free energy of the potential energy minimum. Units: kJ/mol.

The barriers (Table E.3) range from 0.1 to 25.3 kJ/mol in potential energy, and from −0.6 to 26.2 kJ/mol in free energy. The one negative free energy barrier is for conversion from "1 to 8 −" to Point 8. The potential energy barriers for conversion between "8 to 10 (QST3) −", a nearly stacked structure, and "8 to 10 (QST3) +", a T-shaped structure with no stacking, are 17.1 and 2.6 kJ/mol, respectively. The barriers between Point 21 and "13 to 15 −", two rather different extended structures, are 14.6 and 9.7 kJ/mol. The barriers between Point 21 and "15 to 21 −" are 1.1 and 0.1 kJ/mol.

Pair of minima	Barrier (E_a), left-to-right	Barrier (E_a), right-to-left	Barrier (G_a), left-to-right	Barrier (G_a), right-to-left
1-to-6+ ←→ 1-to-6−	2.45	3.56	3.74	3.72
8 ←→ 1-to-8−	24.71	1.30	23.25	−0.60
6-to-7+ ←→ 6-to-7−	4.13	4.29	3.68	6.80
7-to-8+ ←→ 8	18.81	23.90	12.24	26.19
8-to-10 (TS) + ←→ 8-to-10 (TS) −	0.67	3.11	5.82	6.65
8-to-10 (QST3) + ←→ 8-to-10 (QST3) −	17.12	2.61	22.48	5.83
13 ←→ 10-to-13−	16.46	25.30	24.12	25.12
21 ←→ 13-to-15−	14.60	9.71	15.99	14.97
21 ←→ 15-to-21−	1.07	0.11	0.86	2.38

Table E.3. Potential energy (E_a) and free energy (G_a) barriers between some A/T minima (kJ/mol).

E.3.4. A/G

Six distinct minima were obtained from the initial round of optimisations. The lowest SCF energy belongs to Point 4, a stacked structure. Points 11 and 12 could not be optimised due to a close interatomic distance. The attempt to optimise Point 14 exceeded the allowed number of optimisation steps.

The search for transition states yielded four in total. The QST2 search between Points 13 and 21 could not be started, due to an atomic clash in the guess TS generated by Gaussian 09, while the search between Points 1 and 4 experienced convergence issues and was aborted. However, a TS was located by searching between Points 1 and 8. Two of the transition states had a lower SCF energy than the minima they were expected to connect. By "following" the imaginary frequencies of all the transition states, we located a further seven distinct minima. In not a single case did a given pair of these optimisations return both of the expected original minima.

Therefore, in total, 13 A/G minima were located, and they are shown in Figure E.9. The lowest SCF energy belongs to "1 to 8 +", a stacked structure, but the lowest Gibbs free energy is that of Point 8, an intermediate structure between stacked and flipped. The potential energies span 52 kJ/mol while the free energies span 31 kJ/mol.

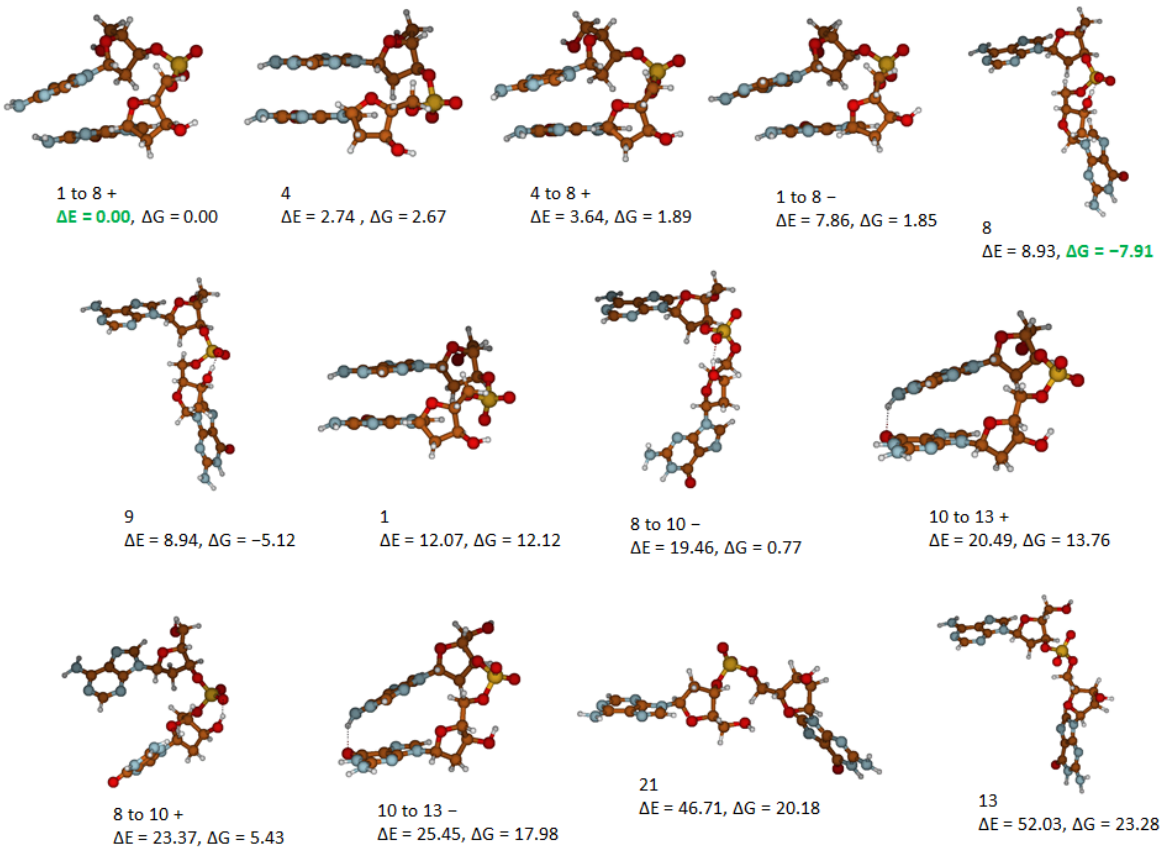


Figure E.9: 13 minima of A/G, ordered by potential energy. Also shown are the free energies, relative to the free energy of the potential energy minimum. Units: kJ/mol.

The barriers (Table E.4) range from 0.1 to 29.9 kJ/mol in potential energy, and from 0.4 to 26.6 kJ/mol in free energy. The potential energy barriers for conversion between "4 to 8 +", a stacked structure, and Point 8, an intermediate structure, are 29.9 and 24.6 kJ/mol. The equivalent barriers between "8 to 10 +" and "8 to 10 −", two rather different intermediate structures, are 5.0 and 8.9 kJ/mol.

Pair of minima	Barrier (E_a), left-to-right	Barrier (E_a), right-to-left	Barrier (G_a), left-to-right	Barrier (G_a), right-to-left
1-to-8 + ←→ 1-to-8 −	7.99	0.13	2.21	0.36
4-to-8 + ←→ 8	29.92	24.64	16.00	25.80
8-to-10 + ←→ 8-to-10 −	5.03	8.94	8.79	13.45
10-to-13 + ←→ 10-to-13 −	22.15	17.18	26.64	22.41

Table E.4. Potential energy (E_a) and free energy (G_a) barriers between some A/T minima (kJ/mol).

E.4. Discussion

It can immediately be seen that, in addition to the stacked and flipped structures, several classes of intermediate structure exist. This is not an entirely new finding. In a DFT study of G/Py dinucleosides (Py = T, U or 5-BrU), using M06-2X as one of the functionals, Churchill and Wetmore identified four structural classes, which they named "distorted", "repelled", "tilted" and "stacked". Many of our structures appear to match these categories. Taking 2AP/T as an example (Figure E.5),

"8 to 9−" (the free energy minimum) is "distorted", Point 7 is "repelled", "7 to 8−" (the potential energy minimum) is "tilted", and Point 1 is "stacked". Churchill and Wetmore did not identify any flipped structures, but in our results these constitute a fifth category (e.g. Point 21 for 2AP/T).

For each of the four systems, the potential energy minimum is either stacked or tilted, while the free energy minimum is distorted. This entropic preference for open structures implies that a considerable fraction of aqueous Pu/Py dinucleosides should be unstacked at room temperature, which may have implications for fluorescence if quenching is stacking-dependent.

The flipped structures labelled 21 in the LIIC – obtained from the experimental geometry of the enzymatically flipped 2AP residue (and adjacent T) in PDB 2IBS – are significantly less stable than the lowest-lying stacked, tilted and distorted structures, in both potential and free energy. In other words, although the entropic drive towards unstacking is strong enough to favour distorted structures, it does not favour fully open ones, at least in the absence of the base flipping enzyme. There appears to be a competition between entropy and stacking, in which distorted structures represent a favourable compromise, while flipped structures are too unstacked. However, we note that Point 21 is not the least stable structure for any system: other unstacked (flipped or distorted) structures of higher energy can always be found.

Despite the general patterns mentioned above, it is clear from Figures E.5–E.9 that structures belonging to the same class are not necessarily similar in energy. In particular, all the dinucleosides have stacked minima that lie quite far above the lowest distorted minima, in potential as well as free energy. In other words, there is considerable energetic variation within as well as between classes.

There are several ways in which the large and diverse set of energy-minimum structures that we have obtained can be geometrically quantified. In Chapter 5, we saw the classification of stacked structures in terms of base step parameters (Twist, Rise, Slide, etc). However, in addition to the fact that the present study considers dinucleosides rather than complete base steps, another problem with applying those parameters here is that they were not designed to describe unstacked structures. It is not clear what meaning the Rise, for example, has in a fully flipped dinucleoside.

An alternative description of DNA structure is given by Svozil *et al.*[10] Here, dinucleosides are analysed in terms of backbone torsions, rather than the relative arrangement of base planes. Each sequence of four backbone atoms, from the C5' deoxyribose carbon of the 5' nucleoside to the O3' deoxyribose oxygen of the 3' nucleoside, defines a torsion. Hence the transition from stacked to flipped configurations of a dinucleoside can be considered an exploration of the backbone conformational space, since the backbone torsions necessarily change as one base is flipped away from the other. This is easier to calculate than base step parameters, since it does not require a coordinate system but is simply a function of the backbone atomic positions. It is also more versatile with respect to structural variation, since the backbone torsions are clearly defined even for flipped geometries, which differ considerably from regular helical structures. Furthermore, torsions are defined independently of one another, unlike base step parameters. We have performed this analysis for each energy minimum and transition state located herein; a comparison with canonical B-DNA values will be presented in a paper that is currently in preparation.

Our group has also carried out a more straightforward structural classification of the energy minima, using just two descriptors: the tilt angle between the planes of each base, and the distance between the centres of each base. This analysis was performed by Rebecca Muller Alcalay. The tilt angles were calculated in the same way as for the stacked dimers in Chapter 5, with the plane of each base defined

by the C2, C4 and C6 atoms. The inter-base distances were calculated by defining the centre of each pyrimidine as the geometric centre of its ring (using the C2, C4 and C6 atoms), and the centre of each purine as the midpoint of its C4-C5 bond. An example plot for 2AP/G structures is given in Figure E.10, showing a roughly linear relationship between tilt angle and inter-base distance. The colour-coded energies also show that the lowest potential energy structures mostly have co-planar base rings that are close together (i.e. stacked), while the highest energy structures mostly have highly tilted base rings that are far apart (i.e. flipped). Between these two extremes, there is considerable energetic variation even among similar structures.

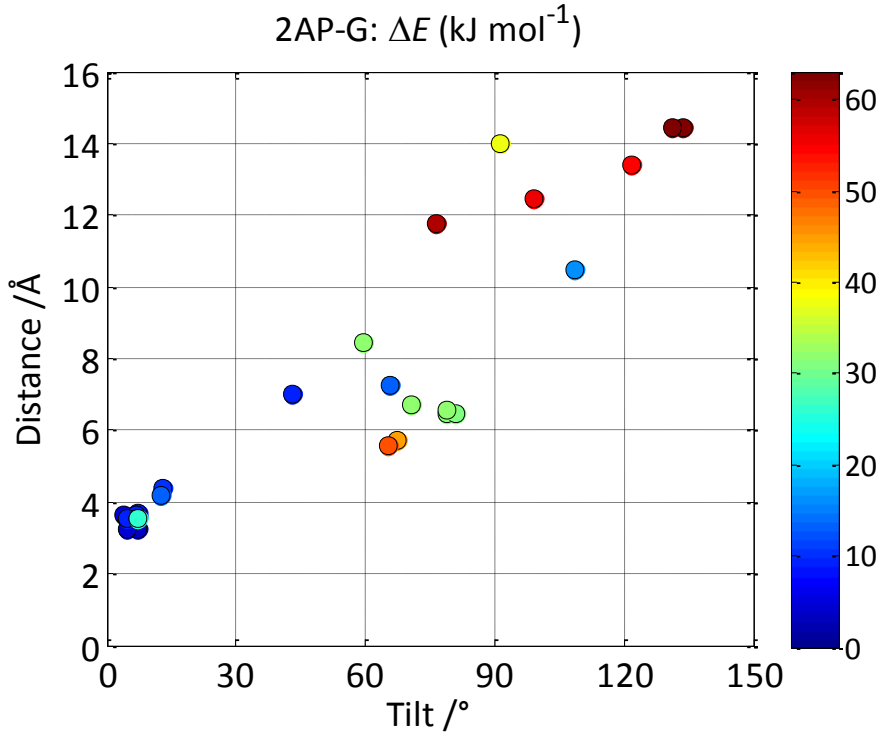


Figure E.10: Scatter plot of angle (Tilt) vs. distance between planes of the 27 2AP/G minima shown in Figure E.7. The colour coding shows the M06-2X potential energies relative to the potential energy minimum structure ("1 to 6–"). Created by Dr. Darren Smith, using data provided by Rebecca Muller Alcalay.

It was disappointing that we could not obtain more information on the energetic barriers separating the minima. However, those barriers which were found (by following the imaginary TS frequencies) to separate structurally distinct minima were large enough (>5 kJ/mol; Tables E.1–E.4) to indicate that the intermediate states between stacked and flipped dinucleosides should correspond to populations of discrete energy minima, rather than a barrier-less continuum. It would be instructive to perform intrinsic reaction coordinate (IRC) calculations on our transition states, to confirm which minima they link.

TD-DFT calculations are now underway in our group to predict the fluorescent properties of the structures discovered in this study, and connect these to the observed fluorescence of 2AP/N dinucleotides.

Appendix F: Publications (including those in preparation)

Published:

"Stacking of the mutagenic DNA base analog 5-bromouracil."
L.F. Holroyd, T. van Mourik, *Theor. Chem. Acc.* **133** (2014) 1431–1443.

"Tyrosine-glycine revisited. Resolving the discrepancy between theory and experiment."
L.F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **621** (2015) 124–128.

In preparation (titles to be confirmed):

"Stacking of the mutagenic DNA base analog 5-bromouracil, Part II. Energy landscapes of pyrimidine dimers in gas phase and water."
L.F. Holroyd, T. van Mourik.

"Comparative stacked structures of dinucleosides containing 2-aminopurine or adenine."
D.A. Smith, A.C. Jones, L.F. Holroyd, T. van Mourik.

"Base flipping pathways of dinucleosides containing 2-aminopurine or adenine."
D.A. Smith, A.C. Jones, L.F. Holroyd, T. van Mourik.

References

[1] J.D. Watson, F.H.C. Crick, Nature 171 (1953) 737-738.
[2] M.H.F. Wilkins, A.R. Stokes, H.R. Wilson, Nature 171 (1953) 738-740.
[3] R.E. Franklin, R.G. Gosling, Nature 171 (1953) 740-741.
[4] J.D. Watson, F.H.C. Crick, Nature 171 (1953) 964-967.
[5] J.D. Watson, F.H.C. Crick, Cold Spring Harb. Symp. Quant. Biol. 18 (1953) 123-131.
[6] H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, Proc. Natl. Acad. Sci. USA 78 (1981) 2179-2183.
[7] G.L. Patrick, An Introduction to Medicinal Chemistry, 2nd ed., Oxford University Press, 2001, ISBN: 198505337.
[8] C.G. Wermuth, C.R. Ganellin, P. Lindberg, L.A. Mitscher, Pure Appl. Chem. 70 (1998) 1129-1143.
[9] G. Elgar, T. Vavouri, Trends Genet. 24 (2008) 344-352.
[10] D. Svozil, J. Kalina, M. Omelka, B. Schneider, Nuc. Acids Res. 36 (2008) 3690-3706.
[11] R.A. Friedman, B. Honig, Biophys. J. 69 (1995) 1528-1535.
[12] K.M. Guckian, B.A. Schweitzer, R.X.-F. Ren, C.J. Sheils, D.C. Tahmassebi, E.T. Kool, J. Am. Chem. Soc. 122 (2000) 2213-2222.
[13] E.T. Kool, Annu. Rev. Biophys. Biomol. Struct. 30 (2001) 1-22.
[14] M. Kabeláč, P. Hobza, Chem. Eur. J. 7 (2001) 2067-2074.
[15] Q.A. Smith, M.S. Gordon, L.V. Slipchenko, J. Phys. Chem. A 115 (2011) 11269-11276.
[16] C.F.R.A.C. Lima, M.A.A. Rocha, L.R. Gomes, J.N. Low, A.M.S. Silva, L.M.N.B.F. Santos, Chem. Eur. J. 18 (2012) 8934-8943.
[17] V.I. Danilov, V.V. Dailidonis, T. van Mourik, H.A. Früchtl, Cent. Eur. J. Chem. 9 (2011) 720-727.
[18] P. Yakovchuk, E. Protozanova, M.D. Frank-Kamenetskii, Nucleic Acids Res. 34 (2006) 564-574.
[19] L.A. Loeb, T.A. Kunkel, Ann. Rev. Biochem. 52 (1982) 429-457.
[20] A. McLennan, A.D. Bates, P. Turner, M. White, BIOS Instant Notes in Molecular Biology, Taylor & Francis, 2012, ISBN: 1136160418, 9781136160417.
[21] M.D. Topal, J.R. Fresco, Nature 263 (1976) 285-289.
[22] J.W. Drake, Proc. Natl. Acad. Sci. USA 88 (1991) 7160-7164.
[23] S.R. Maloy, J.E. Cronan, D. Freifelder, Microbial Genetics, Jones & Bartlett Learning, 1994, ISBN: 0867202483, 9780867202489.
[24] R.H. Tamarin, Principles of Genetics, WCB/McGraw-Hill, 1999, ISBN: 0697354628, 9780697354624.
[25] M. Hanus, M. Kabeláč, J. Rejnek, F. Ryjáček, P. Hobza, J. Phys. Chem. B 108 (2004) 2087-2097.
[26] P.-O. Löwdin, Rev. Mod. Phys. 35 (1963) 724-732.
[27] J. Al-Khalili, Nature's quantum subways, Physics World, Institute of Physics Publishing, 2013, pp. 42-45.
[28] O.O. Brovarets', I.M. Kolomiets', D.M. Hovorun, Elementary Molecular Mechanisms of the Spontaneous Point Mutations in DNA: A Novel Quantum-Chemical Insight into the Classical Understanding, in: T. Tada (Ed.), Quantum Chemistry - Molecules for Innovations, InTech, 2012, ISBN: 9789535103721.
[29] N.K. Sinha, M.D. Haimes, J. Biol. Chem. 256 (1981) 10671-10683.
[30] A.R. Morgan, Trends Biochem. Sci. 18 (1993) 160-163.
[31] F.H.C. Crick, J. Mol. Biol. 19 (1966) 548-555.
[32] T. Brown, O. Kennard, G. Kneale, D. Rabinovich, Nature 315 (1985) 604-606.
[33] W.N. Hunter, T. Brown, G. Kneale, N.N. Anand, D. Rabinovich, O. Kennard, J. Biol. Chem. 262 (1987) 9962-9970.
[34] E. Alexopoulos, E.A. Jares-Erijman, T.M. Jovin, R. Klement, R. Machinek, G.M. Sheldrick, I. Usón, Acta Cryst. D 61 (2005) 407-415.

[35] W.N. Hunter, T. Brown, N.N. Anand, O. Kennard, Nature 320 (1986) 552-555.
[36] W. Wang, H.W. Hellinga, L.S. Beese, Proc. Natl. Acad. Sci. USA 108 (2011) 17644-17648.
[37] P. Strazewski, C. Tamm, Angew. Chem. Int. Ed. Engl. 29 (1990) 36-57.
[38] W. Suen, T.G. Spiro, L.C. Sowers, J.R. Fresco, Proc. Natl. Acad. Sci. USA 96 (1999) 4500-4505.
[39] S. Moran, R.X.-F. Ren, S. Rumney IV, E.T. Kool, J. Am. Chem. Soc. 119 (1997) 2056-2057.
[40] V.H. Harris, C.L. Smith, W.J. Cummins, A.L. Hamilton, H. Adams, M. Dickman, D.P. Hornby, D.M. Williams, J. Mol. Biol. 326 (2003) 1389-1401.
[41] V.H. Harris, C.L. Smith, W.J. Cummins, A.L. Hamilton, D.P. Hornby, D.M. Williams, Org. Biolmol. Chem. 1 (2003) 2070-2074.
[42] Y. Podolyan, L. Gorb, J. Leszczynski, J. Phys. Chem. A 109 (2005) 10445-10450.
[43] P. Khuu, P.S. Ho, Biochem. 48 (2009) 7824-7832.
[44] K. Bebenek, L.C. Pedersen, T.A. Kunkel, Proc. Natl. Acad. Sci. USA 108 (2011) 1862-1867.
[45] B. Schneider, D.M. Cohen, L. Schleifer, A.R. Srinivasan, W.K. Olson, H.M. Berman, Biophys. J. 65 (1993) 2291-2303.
[46] B. Lippert, D. Gupta, Dalton Trans. 24 (2009) 4619-4634.
[47] S.A. Trygubenko, T.V. Bogdan, M. Rueda, M. Orozco, F.J. Luque, J. Šponer, P. Slavíček, P. Hobza, Phys. Chem. Chem. Phys. 4 (2002) 4192-4203.
[48] M. Hanus, F. Ryjáček, M. Kabeláč, T. Kubař, T.V. Bogdan, S.A. Trygubenko, P. Hobza, J. Am. Chem. Soc. 125 (2003) 7678-7688.
[49] J. Rejnek, M. Hanus, M. Kabeláč, F. Ryjáček, P. Hobza, Phys. Chem. Chem. Phys. 7 (2005) 2006-2017.
[50] L. Gorb, Y. Podolyan, P. Dziekonski, W.A. Sokalski, J. Leszczynski, J. Am. Chem. Soc. 126 (2004) 10119-10129.
[51] J. Florián, J. Leszczynski, J. Am. Chem. Soc. 118 (1996) 3010-3017.
[52] R.M. Litman, A.B. Pardee, Nature 178 (1956) 529-531.
[53] S. Benzer, E. Freese, Proc. Natl. Acad. Sci. USA 44 (1958) 112-119.
[54] E. Freese, Proc. Natl. Acad. Sci. USA 45 (1959) 622-633.
[55] T.A. Trautner, M.N. Swartz, A. Kornberg, Proc. Natl. Acad. Sci. USA 48 (1962) 449-455.
[56] B.E. Terzaghi, G. Streisinger, F.W. Stahl, Proc. Natl. Acad. Sci. USA 48 (1962) 1519-1524.
[57] J.W. McGahen, C.E. Hoffmann, Nature 209 (1966) 1241-1242.
[58] J.P. Henderson, J. Byun, D.M. Mueller, J.W. Heinecke, Biochem. 40 (2001) 2052-2059.
[59] M.D. Bick, R.L. Davidson, Proc. Natl. Acad. Sci. USA 71 (1974) 2082-2086.
[60] E.R. Kaufman, R.L. Davidson, Proc. Natl. Acad. Sci. USA 75 (1978) 4982-4986.
[61] C.R. Ashman, R.L. Davidson, Mol. Cell. Biol. 1 (1981) 254-260.
[62] E.R. Kaufman, Mol. Cell. Biol. 4 (1984) 2449-2454.
[63] R.L. Davidson, P. Broeker, C.R. Ashman, Proc. Natl. Acad. Sci. USA 85 (1988) 4406-4410.
[64] R.L. Hopkins, M.F. Goodman, Proc. Natl. Acad. Sci. USA 77 (1980) 1801-1805.
[65] A.R. Katritzky, A.J. Waring, J. Chem. Soc. (1962) 1540-1544.
[66] M. Hanus, M. Kabeláč, D. Nachtigallová, P. Hobza, Biochem. 44 (2005) 1701-1707.
[67] L.A. Allison, Fundamental Molecular Biology, Blackwell Publishing, 2007, ISBN: 9781405103794.
[68] D.L. Hartl, Essential Genetics: A Genomics Perspective, Jones & Bartlett Publishers, 2009, ISBN: 0763782394, 9780763782399.
[69] L. Fishbein, Chemical Mutagens: Environmental Effects on Biological Systems, Elsevier, 2012, ISBN: 0323153259, 9780323153256.
[70] T. Brown, G. Kneale, W.N. Hunter, O. Kennard, Nucleic Acids Res. 14 (1986) 1801-1809.
[71] R. Sanishvili, C. Besnard, F. Camus, M. Fleurant, P. Pattison, G. Bricogne, M. Schiltz, J. Appl. Cryst. 40 (2007) 552-558.
[72] M. Coll, D. Saal, C.A. Frederick, J. Aymami, A. Rich, A.H.-J. Wang, Nucleic Acids Res. 17 (1989) 911-923.
[73] H. Sternglanz, C.E. Bugg, Biochim. Biophys. Acta 378 (1975) 1-11.
[74] N.I. Nakano, S.J. Igarashi, Biochem. 9 (1970) 577-583.

[75] M. Riley, A. Paul, Biochem. 10 (1971) 3819-3825.
[76] L.C. Sowers, R. Eritja, B. Kaplan, M.F. Goodman, G.V. Fazakerley, J. Biol. Chem. 262 (1987) 15436-15442.
[77] C.E. Bugg, J.M. Thomas, M. Sundaralingam, S.T. Rao, Biopolymers 10 (1971) 175-219.
[78] H. Sternglanz, C.E. Bugg, Structural properties of purine and pyrimidine analogs, in: E.D. Bergmann, B. Pullman (Eds.), Molecular and Quantum Pharmacology, D. Reidel Publishing Company, 1974, pp. 460-487, ISBN: 789401017602, 9789401017589.
[79] F.L. Holmes, Meselson, Stahl, and the Replication of DNA: A History of "The Most Beautiful Experiment in Biology", Yale University Press, 2008, ISBN: 0300129661, 9780300129663.
[80] P.H. Driggers, K.L. Beattie, Biochem. 27 (1988) 1729-1735.
[81] H. Yu, R. Eritja, L.B. Bloom, M.F. Goodman, J. Biol. Chem. 268 (1993) 15935-15943.
[82] A.B. Kremer, T. Mikita, G.P. Beardsley, Biochem. 26 (1987) 391-397.
[83] L.C. Sowers, M.F. Goodman, R. Eritja, B. Kaplan, G.V. Fazakerley, J. Mol. Biol. 205 (1989) 437-447.
[84] B. Goz, Pharmacol. Rev. 29 (1978) 249-272.
[85] M. Orozco, B. Hernández, F.J. Luque, J. Phys. Chem. B. 102 (1998) 5228-5233.
[86] X. Hu, H. Li, J. Ding, S. Han, Biochem. 43 (2004) 6361-6369.
[87] V.I. Danilov, T. van Mourik, N. Kurita, H. Wakabayashi, T. Tsukamoto, D. Hovorun, J. Phys. Chem. A 113 (2009) 2233-2235.
[88] T. van Mourik, V.I. Danilov, V.V. Dailidonis, N. Kurita, H. Wakabayashi, T. Tsukamoto, Theor. Chem. Acc. 125 (2010) 233-244.
[89] T. van Mourik, M.-P. Gaigeot, Modelling the dynamics of the keto-enol tautomerisation of uracil and 5-bromouracil using first principles molecular dynamics, HPC-Europa2 report, project no. 228398, 2011.
[90] K. Nomura, R. Hoshino, E. Shimizu, Y. Hoshiba, V.I. Danilov, N. Kurita, J. Mod. Phys. 4 (2013) 422-431.
[91] C.J. Cramer, Essentials of Computational Chemistry: Theories and Models, 2nd ed., John Wiley & Sons, Ltd, 2004, ISBN: 139780470091821.
[92] J.L. McDonagh, Computing the Aqueous Solubility of Organic Drug-Like Molecules and Understanding Hydrophobicity, School of Chemistry, University of St Andrews, St Andrews (UK), 2014.
[93] C.C.J. Roothaan, Rev. Mod. Phys. 23 (1951) 69-89.
[94] G.G. Hall, Proc. R. Soc. A 205 (1951) 541-552.
[95] C. Møller, M.S. Plesset, Phys. Rev. 46 (1934) 618-622.
[96] T. Schwabe, S. Grimme, Acc. Chem. Res. 41 (2008) 569-579.
[97] J. Čížek, J. Chem. Phys. 45 (1966) 4256-4266.
[98] G.D. Purvis III, R.J. Bartlett, J. Chem. Phys. 76 (1982) 1910-1918.
[99] K. Raghavachari, G.W. Trucks, J.A. Pople, M. Head-Gordon, Chem. Phys. Lett. 157 (1989) 479-483.
[100] J. Řezáč, P. Hobza, J. Chem. Theory Comput. 9 (2013) 2151-2155.
[101] P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B 864 - B 871.
[102] W. Kohn, L.J. Sham, Phys. Rev. 140 (1965) A 1133 - A 1138.
[103] P.A.M. Dirac, Proc. Cambridge Philos. Soc. 26 (1930) 376-385.
[104] S.H. Vosko, L. Wilk, M. Nusair, Can. J. Phys. 58 (1980) 1200-1211.
[105] L.H. Thomas, Proc. Cambridge Philos. Soc. 23 (1927) 542-548.
[106] A.D. Becke, Phys. Rev. A 38 (1988) 3098-3100.
[107] A.D. Becke, J. Chem. Phys. 107 (1997) 8554-8560.
[108] C. Adamo, V. Barone, J. Chem. Phys. 108 (1998) 664-675.
[109] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865-3868.
[110] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785-789.
[111] Y. Zhao, N.E. Schultz, D.G. Truhlar, J. Chem. Theory Comput. 2 (2006) 364-382.

[112] Y. Zhao, D.G. Truhlar, Theor. Chem. Acc. 120 (2008) 215-241.
[113] A.D. Becke, J. Chem. Phys. 98 (1993) 5648-5652.
[114] T. Schwabe, S. Grimme, Phys. Chem. Chem. Phys. 8 (2006) 4398-4401.
[115] S.F. Sousa, P.A. Fernandes, M.J. Ramos, J. Phys. Chem. A 111 (2007) 10439-10452.
[116] S. Kozuch, D. Gruzman, J.M.L. Martin, J. Phys. Chem. C. 114 (2010) 20801-20808.
[117] S. Grimme, J. Comput. Chem. 25 (2004) 1463-1473.
[118] S. Grimme, J. Comput. Chem. 27 (2006) 1787-1799.
[119] T. van Mourik, R.J. Gdanitz, J. Chem. Phys. 116 (2002) 9620-9623.
[120] J. Šponer, K.E. Riley, P. Hobza, Phys. Chem. Chem. Phys. 10 (2008) 2595-2610.
[121] T.H. Dunning Jr., J. Chem. Phys. 90 (1989) 1007-1023.
[122] A. Laaksonen, Y. Tu, Methods of incorporating quantum mechanical calculations into molecular dynamics simulations, in: P. Balbuena, J.M. Seminario (Eds.), Molecular Dynamics: From Classical to Quantum Methods, Elsevier, 1999, ISBN: 0080536840, 9780080536842.
[123] S.F. Boys, F. Bernardi, Mol. Phys. 19 (1970) 553-566.
[124] L.F. Holroyd, T. van Mourik, Chem. Phys. Lett. 442 (2007) 42-46.
[125] H. Kruse, L. Goerigk, S. Grimme, J. Org. Chem. 77 (2012) 10824-10834.
[126] R.A. Kendall, H.A. Früchtl, Theor. Chem. Acc. 97 (1997) 158-163.
[127] M.G. Papadopoulos, R. Zalesny, P.G. Mezey, Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications, Springer Science & Business Media, 2011, ISBN: 9048128536, 9789048128532.
[128] H.-J. Werner, F.R. Manby, P.J. Knowles, J. Chem. Phys. 118 (2003) 8149-8160.
[129] M. Schütz, G. Hetzer, H.-J. Werner, J. Chem. Phys. 111 (1999) 5691-5705.
[130] G. Hetzer, P. Pulay, H.-J. Werner, Chem. Phys. Lett. 290 (1998) 143-149.
[131] M. Schütz, J. Chem. Phys. 113 (2000) 9986-10001.
[132] M. Schütz, H.-J. Werner, Chem. Phys. Lett. 318 (2000) 370-378.
[133] H.B. Schlegel, J. Comput. Chem. 3 (1982) 214-218.
[134] J. Tomasi, B. Mennucci, R. Cammi, Chem. Rev. 105 (2005) 2999-3093.
[135] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471-2474.
[136] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford CT, 2009.
[137] A. Furmanchuk, O. Isayev, L. Gorb, O.V. Shishkin, D. Hovorun, J. Leszczynski, Phys. Chem. Chem. Phys. 13 (2011) 4311-4317.
[138] M. Cossi, V. Barone, R. Cammi, J. Tomasi, Chem. Phys. Lett. 255 (1996) 327-335.
[139] A. Klamt, G.J.P. Krooshof, R. Taylor, AIChE J. 48 (2002) 2332-2349.
[140] X. Hu, H. Li, W. Liang, S. Han, J. Phys. Chem. B. 108 (2004) 12999-13007.
[141] M.T. Sykes, M. Levitt, Proc. Natl. Acad. Sci. USA 104 (2007) 12336-12340.
[142] T. Lukmanov, S.P. Ivanov, E.M. Khamitov, S.L. Khursan, Comput. Theor. Chem. 1023 (2013) 38-45.
[143] T. van Mourik, Chem. Phys. 304 (2004) 317-319.
[144] M. Korth, S. Grimme, J. Chem. Theory Comput. 5 (2009) 993-1003.
[145] Y. Zhao, D.G. Truhlar, J. Chem. Phys. 125 (2006) 194101-194118.
[146] C. Adamo, V. Barone, J. Chem. Phys. 110 (1999) 6158-6170.

[147] J. Kästner, J.M. Carr, T.W. Keal, W. Thiel, A. Wander, P. Sherwood, J. Phys. Chem. A 113 (2009) 11856-11865.
[148] ChemShell, a Computational Chemistry Shell, see www.chemshell.org.
[149] N. Mardirossian, M. Head-Gordon, J. Chem. Theory Comput. 9 (2013) 4453-4461.
[150] P. Auffinger, F.A. Hays, E. Westhof, P.S. Ho, Proc. Natl. Acad. Sci. USA 101 (2004) 16789-16794.
[151] K.E. Riley, P. Hobza, Phys. Chem. Chem. Phys. 15 (2013) 17742-17751.
[152] M. Kolář, P. Hobza, A.K. Bronowska, Chem. Commun. 49 (2013) 981-983.
[153] V.K. Rastogi, M.A. Palafox, L. Mittal, N. Peica, W. Kiefer, K. Lang, S.P. Ojha, J. Raman Spectrosc. 38 (2007) 1227-1241.
[154] R.S. Lasken, M.F. Goodman, J. Biol. Chem. 259 (1984) 11491-11495.
[155] B. Glickman, P. van den Elsen, M. Radman, Molec. Gen. Genet. 163 (1978) 307-312.
[156] G. Fogarasi, Chem. Phys. 349 (2008) 204-209.
[157] L. Gorb, Y. Podolyan, J. Leszczynski, W. Siebrand, A. Fernández-Ramos, Z. Smedarchina, Biopolymers 61 (2002) 77-83.
[158] N. Markova, V. Enchev, I. Timtcheva, J. Phys. Chem. A 109 (2005) 1981-1988.
[159] M. Sprik, G. Ciccotti, J. Chem. Phys. 109 (1998) 7737-7744.
[160] I. Ivanov, M.L. Klein, J. Am. Chem. Soc. 124 (2002) 13380-13381.
[161] T. van Mourik, M.-P. Gaigeot, personal communication, 2012.
[162] Grace: http://plasma-gate.weizmann.ac.il/Grace/.
[163] CPMD, copyright International Business Machines Corporation (1990-2008) and Max Planck Institute fuer Festkoerperforschung Stuttgart (1995-2001), pp. CPMD, copyright International Business Machines Corporation (1990-2008) and Max Planck Institute fuer Festkoerperforschung Stuttgart (1995-2001).
[164] A.D. Becke, J. Chem. Phys. 96 (1992) 2155-2159.
[165] N. Troullier, J.L. Martins, Phys. Rev. B 43 (1991) 1993-2006.
[166] L. Kleinman, D.M. Bylander, Phys. Rev. Lett. 48 (1982) 1425-1428.
[167] W. Humphrey, A. Dalke, K. Schulten, J. Molec. Graphics 14 (1996) 33-38.
[168] VMD - (V)isual (M)olecular (D)ynamics: http://www.ks.uiuc.edu/Research/vmd/.
[169] M. Bühl, personal communication, 2014-2015.
[170] E. Kimura, H. Kitamura, T. Koike, M. Shiro, J. Am. Chem. Soc. 119 (1997) 10909-10919.
[171] M.A. Kurinovich, J.K. Lee, J. Am. Soc. Mass Spectrom. 13 (2002) 985-995.
[172] Y.H. Jang, L.C. Sowers, T. Çağin, W.A. Goddard III, J. Phys. Chem. A 2001 (2001) 274-280.
[173] M.-P. Gaigeot, A. Cimas, M. Seydou, J.-Y. Kim, S. Lee, J.P. Schermann, J. Am. Chem. Soc. 132 (2010) 18067-18077.
[174] M. Sprik, Chem. Phys. 258 (2000) 139-150.
[175] H.M. Senn, S. Thiel, W. Thiel, J. Chem. Theory Comput. 1 (2005) 494-505.
[176] P. López-Tarifa, M.-P. Gaigeot, R. Vuilleumier, I. Tavernelli, M. Alcamí, F. Martín, M.-A.H. du Penhoat, M.-F. Politis, Angew. Chem. Int. Ed. 52 (2013) 3160-3163.
[177] S. Izvekov, G.A. Voth, J. Chem. Phys. 123 (2005) 044505-044509.
[178] T. Todorova, A.P. Seitsonen, J. Hutter, I.-F.W. Kuo, C.J. Mundy, J. Phys. Chem. B 110 (2006) 3685-3691.
[179] M.A. El Hassan, C.R. Calladine, J. Mol. Biol. 251 (1995) 648-664.
[180] R.E. Dickerson, M. Bansal, C.R. Calladine, S. Diekmann, W.N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H.C.M. Nelson, W.K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D.M. Soumpasis, C.-S. Tung, A.H.-J. Wang, V.B. Zhurkin, EMBO J. 8 (1989) 1-4.
[181] X.-J. Lu, W.K. Olson, Nucleic Acids Res. 31 (2003) 5108-5121.
[182] W.K. Olson, M. Bansal, S.K. Burley, R.E. Dickerson, M. Gerstein, S.C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger, H.M. Berman, J. Mol. Biol. 313 (2001) 229-237.
[183] M.A. El Hassan, C.R. Calladine, Phil. Trans. R. Soc. Lond. A 355 (1997) 43-100.
[184] J.M. Sturtevant, S.A. Rice, E.P. Geiduschek, Disc. Faraday Soc. 25 (1958) 138-149.

[185] P.O.P. Ts'o, S.I. Chan, J. Am. Chem. Soc. 86 (1964) 4176-4181.
[186] S.H. Gellman, T.S. Haque, L.F. Newcomb, Biophys. J. 71 (1996) 3523-3525.
[187] R.A. Friedman, B. Honig, Biophys. J. 71 (1996) 3525-3526.
[188] R. Luo, H.S.R. Gilson, M.J. Potter, M.K. Gilson, Biophys. J. 80 (2001) 140-148.
[189] J. Florián, J. Šponer, A. Warshel, J. Phys. Chem. B 103 (1999) 884-892.
[190] P. Jurečka, P. Hobza, J. Am. Chem. Soc. 125 (2003) 15608-15613.
[191] I. Dąbkowska, H.V. Gonzalez, P. Jurečka, P. Hobza, J. Phys. Chem. A 109 (2005) 1131-1136.
[192] J. Šponer, P. Jurečka, I. Marchan, F.J. Luque, M. Orozco, P. Hobza, Chem. Eur. J. 12 (2006) 2854-2865.
[193] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A.K. Wilson, Chem. Phys. Lett. 286 (1998) 243-252.
[194] Ż. Czyżnikowska, R. Zaleśny, M. Ziółkowski, R.W. Gora, P. Cysewski, Chem. Phys. Lett. 450 (2007) 132-137.
[195] V.R. Cooper, T. Thonhauser, A. Puzder, E. Schröder, B.I. Lundqvist, D.C. Langreth, J. Am. Chem. Soc. 130 (2007) 1304-1308.
[196] C.A. Morgado, P. Jurečka, D. Svozil, P. Hobza, J. Šponer, J. Chem. Theory Comput. 5 (2009) 1524-1544.
[197] J. Trnka, R. Sedlak, M. Kolář, P. Hobza, J. Phys. Chem. A 117 (2013) 4331-4337.
[198] J. Šponer, J.E. Šponer, A. Mládek, P. Jurečka, P. Banás, M. Otyepka, Biopolymers 99 (2013) 978-988.
[199] J. Iball, C.H. Morgan, H.R. Wilson, Proc. R. Soc. Lond. A 295 (1966) 320-333.
[200] J. Petruska, L.C. Sowers, M.F. Goodman, Proc. Natl. Acad. Sci. USA 83 (1986) 1559-1562.
[201] RCSB Protein Data Bank, www.pdb.org.
[202] M.S. Marshall, L.A. Burns, C.D. Sherrill, J. Chem. Phys. 135 (2011) 194102-194110.
[203] F.R. Manby, personal communication, 2012.
[204] H.-J. Werner, P.J. Knowles, G. Knizia, F.R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K.R. Shamasundar, T.B. Adler, R.D. Amos, A. Bernhardsson, A. Berning, D.L. Cooper, M.J.O. Deegan, A.J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A.W. Lloyd, R.A. Mata, A.J. May, S.J. McNicholas, W. Meyer, M.E. Mura, A. Nicklass, D.P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A.J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, MOLPRO, version 2010.1, a package of ab initio programs, see http://www.molpro.net.
[205] F. Neese, The ORCA program system, WIREs Comput. Mol. Sci. 2 (2012) 73.
[206] R.S. Hunter, T. van Mourik, J. Comput. Chem. 33 (2012) 2161-2172.
[207] Y. Li, G. Zon, W.D. Wilson, Proc. Natl. Acad. Sci. USA 88 (1991) 26-30.
[208] S.-H. Chou, Y.-Y. Tseng, J. Mol. Biol. 285 (1999) 41-48.
[209] MOLDEN: a pre- and post processing program of molecular and electronic structure, Version 5.0, See; http://www.cmbi.ru.nl/molden/, 2010.
[210] C. Peng, H.B. Schlegel, Israel J. Chem. 33 (1993) 449-454.
[211] L.F. Holroyd, T. van Mourik, Theor. Chem. Acc. 133 (2014) 1431-1443.
[212] T. van Mourik, personal communication, 2013.
[213] M. Nakata, G. Zanchetta, B.D. Chapman, C.D. Jones, J.O. Cross, R. Pindak, T. Bellini, N.A. Clark, Science 318 (2007) 1276-1279.
[214] J.-D. Chai, M. Head-Gordon, Phys. Chem. Chem. Phys. 10 (2008) 6615-6620.
[215] D. Shugar, B. Kierdaszuk, Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. 8 (1985) 657-668.
[216] N. Kurita, V.I. Danilov, V.M. Anisimov, D. Hovorun, T. Nakatsu, K. Dedachi, Mechanisms of Mutagenic Action of 5BrU: Quantum Mechanical Study, pp. 665.
[217] L.C. Sowers, R. Eritja, B. Kaplan, M.F. Goodman, G.V. Fazakerly, J. Biol. Chem. 263 (1988) 14794-14801.

[218] J.A. Theruvathu, C.H. Kim, A. Darwanto, J.W. Neidigh, L.C. Sowers, Biochem. 48 (2009) 11312-11318.
[219] J.M. Vargason, B.F. Eichman, P.S. Ho, Nat. Struct. Biol. 7 (2000) 758-761.
[220] D.B. Longley, D.P. Harkin, P.G. Johnston, Nat. Rev. Cancer 3 (2003) 330-338.
[221] A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma, N.P. Ernsting, Angew. Chem. Int. Ed. Engl. 49 (2010) 5989-5992.
[222] O.J.G. Somsen, A. van Hoek, H. van Amerongen, Chem. Phys. Lett. 402 (2005) 61-65.
[223] T. Lenz, E.Y.M. Bonnist, G. Pljevaljčić, R.K. Neely, D.T.F. Dryden, A.J. Scheidig, A.C. Jones, E. Weinhold, J. Am. Chem. Soc. 129 (2007) 6240-6248.
[224] O.F.A. Larsen, I.H.M. van Stokkum, F.L. de Weerd, M. Vengris, C.T. Aravindakumar, R. van Grondelle, N.E. Geacintov, H. van Amerongen, Phys. Chem. Chem. Phys. 6 (2004) 154-160.
[225] J. Liang, S. Matsika, J. Am. Chem. Soc. 133 (2011) 6799-6808.
[226] E.Y.M. Bonnist, A.C. Jones, Chem. Phys. Chem. 9 (2008) 1121-1129.
[227] R.K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys, A.C. Jones, Nuc. Acids Res. 37 (2009) 6859-6870.
[228] J.M. Jean, K.B. Hall, Proc. Natl. Acad. Sci. USA 98 (2001) 37-41.
[229] J. Leszczynski, Handbook of Computational Chemistry, Springer Science & Business Media, 2012, ISBN: 9400707118, 9789400707115.
[230] V.I. Poltev, V.M. Anisimov, V.I. Danilov, T. van Mourik, A. Deriabina, E. Gonzalez, M. Padua, D. Garcia, F. Rivas, N. Polteva, Int. J. Quant. Chem. 110 (2010) 2548-2559.
[231] V. Poltev, V.M. Anisimov, V.I. Danilov, D. Garcia, A. Deriabina, E. Gonzalez, R. Salazar, F. Rivas, N. Polteva, Comput. Theor. Chem. 975 (2011) 69-75.
[232] C.D.M. Churchill, S.D. Wetmore, Phys. Chem. Chem. Phys. 13 (2011) 16373-16383.
[233] V. Poltev, V.M. Anisimov, V.I. Danilov, D. Garcia, C. Sanchez, A. Deriabina, E. Gonzalez, F. Rivas, N. Polteva, Biopolymers 101 (2013) 640-650.